NONLINEAR SINGULAR SPECTRUM ANALYSIS AND ITS APPLICATION IN MULTIVARIATE STATISTICAL PROCESS MONITORING

by

Syamala Krishnannair

Dissertation presented for the degree

of

DOCTOR OF PHILOSOPHY
(EXTRACTIVE METALLURGICAL ENGINEERING)

in the Faculty of Engineering
at Stellenbosch University

Supervisor
Prof. C. Aldrich

Co-Supervisor
Prof. S.M. Bradshaw

March 2016

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2016

SUMMARY

Multivariate statistical process control (MSPC) approaches based on principal component analysis (PCA), partial least squares (PLS) and related extensions are now widely used for process monitoring and diagnosis in process systems where observed correlated measurements are readily available. However, highly nonlinear (dynamic) processes pose a challenge for MSPC methods as a large set of nonlinear features are typically required to capture the underlying characteristic behaviour of the process in the absence of faults. Several extensions of basic (PCA) methods have previously been proposed to handle features such as autocorrelation in data, time-frequency localization, and nonlinearity.

In this study multivariate statistical process monitoring methods based on nonlinear singular spectrum analysis which use nonlinear principal component analysis, multidimensional scaling and kernel multidimensional scaling are proposed. More specifically, singular spectrum analysis using covariance and dissimilarity scale structure are proposed to express multivariate time series as the sum of identifiable components whose basis functions are obtained from the process measurements. Such an approach is useful for extracting trends, harmonic patterns and noise in time series data. Using nonlinear SSA decomposition of time series data, a multimodal representation is obtained that can be used together with existing statistical process control methods to develop novel process monitoring schemes.

The advantages of these approaches are demonstrated on simulated multivariate nonlinear data and compared with those of classical PCA and multimodal SSA on base metal flotation plant data and the Tennessee Eastman process benchmark data. The nonlinear SSA methods better captured the nonlinearities in the observed data. Consequently, this yielded improved detection rates for various faults in nonlinear data over those obtainable by alternative competing multivariate methods.

OPSOMMING

Meerveranderlike statistiese prosesbeheer (MSP) benaderings gebaseer op hoofkomponentontleding, gedeeltelike kleinste kwadrate en verwante uitbreidings, word tans wyd gebruik in prosesmonitering en –diagnose van prosesstelsels waar waargenome gekorreleerde metings geredelik beskikbaar is. Hoogs nie-lineêre (dinamiese) prosesse is egter ’n uitdaging vir MSP metodes, aangesien ’n groot stel nie-lineêre kenmerke tipies benodig word om die onderliggende karakteristieke gedrag van die proses vas te vang in die afwesigheid van foute. Verskeie uitbreidings van basiese hoofkomponentonledingsmetodes is voorheen voorgestel om kenmerke, soos outokorrelasie, tyd-frekwensielokalisering en nie-lineariteit in die data te hanteer.

In die studie, word meerveranderlike statistiese prosesmoniteringsmetodes voorgestel, gebaseer op nie-lineêre enkelvoudige spektrumontleding wat nie-lineêre hoofkomponentontleding, meerdimensionele skalering en kern- multidimensionele skalering gebruik. Meer spesifiek, enkelvoudige spektrumontleding wat kovariansie- en andersheidskaalstrukture gebruik, word voorgestel om meerveranderlike tydreekse uit te druk as die som van identifiseerbare komponente, wat se basisfunksies van prosesmetings verkry kan word. So ’n benadering is nuttig vir die ekstraksie van tendense, harmoniese patrone en geraas in die tydreeksdata. Deur nie-lineêre enkelvoudige spektrumontleding te gebruik vir ontbinding van die tydreeksdata, word ’n multimodale verteenwoordiging verkry wat gebruik kan word saam met bestaande statistiese prosesbeheermetodes om nuwe prosesmoniteringskemas te ontwikkel.

Die voordele van die benaderings word gedemonstreer en vergelyk met die van klassieke hoofkomponentontleding en multimodale nie-lineêre enkelvoudige spektrumontleding op gesimuleerde meerveranderlike nie-lineêre data, data van ’n basismetaalflottasie-aanleg en die Tennessee Eastman prosesykingsdata. Die nie-lineêre enkelvoudige spektrumontledingsmetodes het die nie-lineariteite in die waargenome prosesdata beter beskryf. Gevolglik het dit tot beter foutopsporingstempo’s gelei, as wat behaal kon word met alternatiewe kompterende meerveranderlike metodes.

ACKNOWLEDGEMENTS

I hereby express my sincere gratitude towards my supervisors, Prof.Chris Aldrich and Prof.Steven Bradshaw for their support and guidance for the successful completion of this study. I am also grateful to Dr.Gorden Jemwa for his help and encouragement.

I would like to thank the administrive staff of the Department of Process Engineering for their prompt help in all the administrative work related to the completion of this study.I would also like to thank University of Zululand for granting me study leave for the completion of the study. I am also very much grateful to my colleagues and friends at University of Zululand for their support.

Finally, I would like to acknowledge the encouragement and support from my husband and my two children which inspired me tremendously.

TABLE OF CONTENTS

ABBREVIATIONS

ANN	Artificial Neural Network
ANOC	Abnormal Operating Condition
CMDS	Classical Multidimensional Scaling
CUSUM	Cumulative sum
DSSA	Singular Spectrum Analysis with Dissimilarity Matrix
EOF	Empirical Orthogonal Function
EWMA	Exponentially Weighted Moving Average
KDSSA	Singular Spectrum Analysis with Kernel Multidimensional Scaling
KMDS	Kernel Multidimensional Scaling
KPCA	Kernel Principal Component Analysis
MCUSUM	Multivariate Cumulative Sum
MEWMA	Multivariate Exponentially Weighted Moving Average
MDS	Multidimensional Scaling
MLP	Multilayer Perception
MM-DSSA	Multimodel Singular Spectrum Analysis with Dissimilarity Matrix
MM-KDSSA	Multimodel Singular Spectrum Analysis with Kernel Multidimensional Scaling
MM-NLSSA	Multimodel Nonlinear Singular Spectrum Analysis
MRA	Multiresolution Analysis
MS-SSA	Multiscale Singular Spectrum Analysis
MSPC	Multivariate Statistical Process Control
NLPCA	Nonlinear Principal Component Analysis
NOC	Normal Operating Condition
PC	Principal Component
PCs	Principal Components
PCA	Principal Component Analysis
PLS	Partial Least Squares
SPC	Statistical Process Control
SPE	Squared Prediction Errors
SSA	Singular Spectrum Analysis

SVD	Singular Value Decomposition

NOMENCLATURE

$\mathbf{a}_i^*$	i^{th} Original eigenvector in CMDS
$\mathbf{a}_i$	i^{th} Normalized eigenvector in CMDS
$\mathbf{B}$	Scalar product matrix
$\mathbf{b}^{(\mathbf{x})}$	Vector of bias parameter
$\mathbf{b}_{ij}$	The scalar product between variables
$\mathbf{C}$	Variance-covariance matrix
$\mathbf{C}^f$	Covariance matrix in feature space
$\mathbf{C}_\mathbf{x}$	Covariance matrix of the trajectory matrix
$c_i(M)$	Autocorrelation function of i^{th} variable
$\mathbf{D}$	Dissimilarity matrix
d_{ij}^2	Squared Euclidean distance between variables
$\mathbf{E}$	Residual matrix
$\mathbf{e}_i$	i^{th} Row of residual matrix
$\tilde{\mathbf{e}}_i$	Normalized eigenvector in kernel MDS
$\tilde{\mathbf{E}}$	The residual matrix
$\tilde{\mathbf{e}}_k$	The k^{th} row of $\tilde{\mathbf{E}}$
$F_{k,n-k,\alpha}$	F Distributional index with k and $n-k$ degrees of freedom at α confident limit
$\boldsymbol{F}$	De-mapping function in NLPCA
$\mathcal{F}$	Feature space in KPCA
$\mathbf{F}$	Input space in kernel MDS
$\mathbf{f}_i$	Input vector in kernel MDS
$\boldsymbol{G}$	Mapping function in NLPCA
$\mathbf{I}$	Identity matrix
$\mathbf{I}_n$	Identity matrix of size $n \times n$
$\mathbf{J}_n$	Centering matrix of size $n \times n$
k	Number of PCs retained in the PCA model
$\boldsymbol{K}_{ij}$	Kernel matrix
$\boldsymbol{k}(\mathbf{x_i}, \mathbf{x}_j)$	Kernel function evaluated on vectors $\mathbf{x}_i$and $\mathbf{x}_j$

K	Length of lagged vectors
M	Embedding window length
m	Number of process variables
n	Number of observations
$\tilde{\mathbf{p}}_{ij}$	j^{th} Eigenvector of i^{th} representation of original Data
$\boldsymbol{Q}$	Squared prediction error statistics
$\widetilde{\boldsymbol{Q}}_k$	$\boldsymbol{Q}$ Statistic for sample k
R_k	Percentage contribution of an eigenvector
$r(.)$	Diagonal averaging function
$r(\mathbf{t}_k, \mathbf{p}_{k,i}^T)$	k^{th}Reconstructed component of k^{th} variable in the i^{th} mode
$\mathbf{T}_k$	Matrix of retained score vectors in PCA
$\mathbf{t}_i$	i^{th} Score vector
$\boldsymbol{T}^2$	Hotelling T square statistic
$\boldsymbol{T}_i^2$	i^{th} Row of T square value
$\mathbf{t}_{ij}$	PC in the i^{th} row and j^{th} column
$\mathbf{t}_{i,new}$	i^{th} Row of new scores
$\mathbf{t}_k(t)$	Principal component scores of dimension K
$\mathbf{T}$	Score matrix
$\mathbf{t}(t)$	Score vectors of dimension $p < M$
$\tilde{\mathbf{t}}_{i,j}$	j^{th}Score of i^{th} representation of original data
$\boldsymbol{T}_k^2$	Hotelling's T^2 statistic for sample k
$\widetilde{\mathbf{T}}_A$	The matrix of A score vectors
$\tilde{\mathbf{t}}_k$	k^{th} Row of $\widetilde{\mathbf{T}}_A$
$\mathbf{U}$	Eigenvector matrix for PCA
$\mathbf{V}_k$	Matrix of retained eigenvectors in PCA
$\mathbf{V}_i$	i^{th} Eigenvector
$\mathbf{V}^{(\mathbf{x})}$	Weight matrix
$\mathbf{v}$	Eigenvector in the feature space
$\mathbf{w}$	Projection axis in PCA
$\mathbf{X}$	Original data matrix
$\widehat{\mathbf{X}}$	Approximated data matrix in PCA
x_i	i^{th} Observation in the original data matrix

$\hat{x}_{ij}$	Approximated value of x_{ij}
$\mathbf{X}_{new}$	New set of observation matrix
$\widehat{\boldsymbol{X}}_{new}$	Approximated matrix of $\mathbf{X}_{new}$
$\widetilde{\mathbf{X}}$	Approximated matrix of the original matrix $\mathbf{X}$
$\tilde{\mathbf{x}}$	Approximated vector of the original vector $\mathbf{x}$
$x(t)$	Time series of length N in SSA
$\mathbf{X}_i$	Lagged vectors of length K
x_{ij}	Elements of trajectory matrix at row 'i' and column 'j'
$\tilde{\mathbf{x}}(t+j-1)$	Elementary matrix of dimension $K \times p$
$\tilde{\mathbf{x}}(i)$	Reconstructed component of original time Series
$\mathbf{X}'_k$	Trajectory matrix of k^{th} variable
$\mathbf{X}_i(l)$	i^{th} Element of variable l
$\overline{\mathbf{X}}_i$	Arithmetic mean of i^{th} variable
$\widetilde{\mathbf{X}}_i$	Reconstructed variable at i^{th} mode
α_{adj}	Adjusted significance level
$\theta_{threshold}$	Percentage of variance captured in PCA
$\boldsymbol{\Lambda}_p$	Diagonal matrix with p eigenvalues
$\boldsymbol{\lambda}$	Diagonal matrix of Eigenvalues for PCA
λ_i	i^{th} Eigenvalue
$\lambda_{i,j}$	j^{th} Eigenvector of i^{th} representation of original data
$\phi(\mathbf{x}_j)$	Nonlinear mapping of the input vector $\mathbf{x}_j$
$\mathbf{1}_n$	Vector with elements of 1 and size n
χ^2	Chi-square

CHAPTER 1: INTRODUCTION

In this chapter, the background and motivation for data-based process monitoring methods for fault or hazard identification and risk management to ensure optimal process efficiencies as well as the safety of operating such chemical processes is introduced.After discussing limitations of current and well-established statistical approaches now routinely used in many industries, the problem statement and objectives underlying this study are presented.

1.1 BACKGROUND

Process monitoring, also referred to as fault detection and diagnosis, is a critical task in any production environment that affects process and product quality, process safety, productivity, material and energy consumption, and therefore the bottom line – that is, the profitability of operating any process plant. Modern chemical and metallurgical processes are characterised by complex physico-chemical phenomena, with interdependence among different unit operations arising from, for example, recycle streams, the effect of feedback control systems and buffer holdups within the system. The features of these processes are highly nonlinear, with long time delays, and highy correlation among the measurements. Hence, ensuring that the operations satisfy the performance specifications and that anomalies are detected at the earliest possible time – which are the objectives of process monitoring – has become a challenge in most practical situations (Li & Xiao, 2011).

Process monitoring approaches are generally divided into the following categories: model-based methods, knowledge-based methods and data-based methods. Model-based methods rely on the physical and mathematical knowledge of the industrial processes and give more accurate results than other two methods, as long as there is *a priori* knowledge of the first principle of physical/chemical relationship between the variables in the process. However, it is difficult to construct first-principles models based on prior knowledge of the process, especially for the modern complex process system.

Knowledge-based methods are built on the available knowledge of the process characteristics and experience of the plant operators gained from oerating the process

over a long period. The results thus obtained by these methods are conclusions that are derived from known facts. However, the developments of knowledge-based models require long-term knowledge and experience of the process operations, and this became a difficult and time-consuming task for process operators. The major difficulty associated with modern chemical process monitoring is the lack of sufficient knowledge to develop accurate mathematical process descriptions, while inappropriate methodologies are used to exploit abundant operational data (Frank, 1990; Ge *et al*., 2013; Himmelblau, 1978; Li & Xiao, 2011).

Recently, data-based process monitoring methods have gained a lot of attention in the monitoring of complex process systems. This is due to the advances in automation technologies, and the large amounts of data being recorded and collected, which contain most of the process information for modelling, monitoring and control (Ge *et al*., 2013). In the chemical industry, for example, an integrated database is used to store all operational data from all process units and also product analysis data obtained by inspection. In such cases, data-based monitoring method can be utilised to extract information from those datasets. Data-based approaches exploit structures or regularities in data to derive mathematical or statistical models that describe expected process behaviour under normal operating conditions. The developed models can then be used for monitoring, control and process optimisation tasks.

Product quality improvement is mainly coupled with tasks such as

(i) prediction of the quality of the product from operating conditions,

(ii) development of better operating conditions that can maintain and improve the product quality and, finally,

(iii) detection and elimination of process faults or malfunctions to prevent any abnormal operations.

Typically, task (i) is implemented by using a soft sensor, in which a mathematical model is built to relate the product quality based on the operating conditions and parameters of the process. This model is then used to formulate and solve an optimisation problem for the realisation of task (ii). The third task is realised by statistical process control (SPC) techniques. SPC is based on the use of probability theory and statistical methods to detect the existence and time of occurrence of faults that cause abnormalities in process performance (Negiz & Cinar, 1997). The SPC framework was originally developed for

industrial engineering applications to monitor stochastic processes. Later, the basic framework of SPC was extended to monitor dynamic multivariate systems with highly correlated variables. The group of such multivariate statistical techniques is commonly referred to as *multivariate statistical process control (MSPC)* (Ge *et al*., 2013; Kano & Nakagawa, 2008; Kresta *et al*., 1991; Krishnannair, 2010; Yin *et al*., 2014).

1.2 MOTIVATION

MSPC techniques focus on exploiting a high degree of redundancy in correlated data by generating a reduced set of statistically uncorrelated variables. Thus the use of multivariate data analysis techniques in process monitoring is easier and less complicated because it provides a reduced variable space in which variables are uncorrelated. Latent variable methods, such as principal component (PC) analysis (PCA) and partial least squares (PLS) are widely used in MSPC in various industrial applications.

Despite the huge success in the application of latent variable methods, it is generally accepted that chemical processes are nonlinear, and the use of methods based on linear correlations in the observed data has the undesirable effect of substantial loss of information (Maulud *et al*., 2006). Therefore, their application to data from a (highly) nonlinear process may result in misleading insights. The application of conventional MSPC techniques to nonlinear data causes difficulties in the selection of an optimal number of latent variables, since the nonlinear effects tend to distribute non-uniformly among the latent variables, that is, latent variables judged uninformative may actually have greater impact.

To overcome the limitations of linear MSPC techniques, several nonlinear extensions of MSPC have been developed to improve feature extraction in the nonlinear data. These include nonlinear generalisation of conventional PCA with auto-associative neural networks, kernel PCA (KPCA) by combining kernel functions and PCA, and multidimensional scaling (MDS) (Cox & Cox, 1994; Dong & McAvoy, 1996; Kramer, 1991; Schölkopf *et al.*, 1998). A detailed review of some of these methods is presented in the next chapter.

MSPC techniques are best at decorrelating variables. However, serial or auto-correlation in individual variables presents a challenge. Auto-correlation in variables can arise from,

for example, the effect of feedback control systems, time correlation of process noises and the inherent dynamic nature of the process. One way to eliminate the dynamic effect of process data is to use a large sampling interval, which will weaken the correlations between the sampling data. However, this leads to a potential loss of information, as well as distortion of the relationships between process variables. In particular, the monitoring performance of the dynamic change in the process variables will deteriorate if the dynamic relationships are not accounted for in process modelling. Hence, several extensions of MSPC have been proposed to handle the static and dynamic process characteristics of the chemical process system (Ge *et al.,* 2013; Ku *et al*., 1995).

The operating conditions of chemical processes vary frequently, mainly because of the fluctuations of process raw materials, the shift of set-points, ageing of process units and seasonal effects, among other factors. MSPC models rely on the assumption that the process operates in a steady-state condition and have practical limitations when applied to chemical processes that are slow-varying or have multiple operating conditions. Moreover, these techniques are based on a single-scale representation of the measurements and hence do not make use of the advantage of multiscale representations of measurements, that is, the representation of features that occur at different localization in time, space and frequency (Aradhye *et al.*, 2003). The decomposition of data into different levels exploits the features of the process that change in frequency and scale. For example, a low-scale (high-frequency) component of a signal gives a detailed view of the signal, while a high-scale (low-frequency) signal component provides a non-detailed view of the signal (Polikar, 1996). An alternative to using linear time series models to capture the dynamics is multiscale modelling using wavelet analysis (Aradhye *et al.*, 2003; Bakshi, 1998; Yoon & MacGregor, 2004). A multiscale approach to statistical process monitoring that combines the PCA and multiscale representation using wavelets was introduced by Bakshi (1998).

In the last decade, singular spectrum analysis (SSA) has been proposed for pre-processing data prior to the application of MSPC in process monitoring (Aldrich & Barkhuizen, 2003; Jemwa & Aldrich, 2006; Krishnannair, 2010). Using SSA, a method for process monitoring called *multiscale SSA* (MS-SSA) was proposed as an alternative to wavelet decomposition for extracting trends at different resolutions. (Although multiscale SSA shares a similar design concept to multiresolution analysis (MRA) in that

a signal is represented at different levels of granularity, the notion of scale as used in MRA is lacking in SSA. In particular, “singular spectrum analysis” relates to the spectrum of eigenvalues from the SVD of the correlation matrix, and not time-frequency decomposition as used in wavelet analysis.) Multiscale SSA allows for simultaneous extraction of complex trends and periodicities with varying amplitude in the process data at different resolutions (Krishnannair, 2010). In this approach, SSA is used to decompose the data into multiple modes/levels, and each signal in the multimodal setup is monitored separately using PCA for fault detection (Krishnannair, 2010). In this method, the word “*scale*” is used to represent the decomposed SSA components at the same mode for different measurements to extract information on the dynamical system. In SSA, the additive components, such as trend, anharmonic oscillations or noise, are extracted by an adaptive basis function that is derived from the data and hence is more appropriate for pre-processing nonlinear and nonstationary data than wavelets, which use a non-adaptive basis function. For a time series with a complex structure, SSA decomposition can be modified in different ways, leading to different versions, such as SSA with single and double centring, toeplitz SSA, sequential SSA, oblique SSA, deriv SSA, kernel SSA and nonlinear SSA (Golyandina & Shlemov, 2014; Golyandina *et al*., 2001; Hsieh & Wu, 2002; Jemwa & Aldrich, 2006). This observation can be exploited, as proposed in this thesis, such that, in the SSA decomposition stage, a dissimilarity scale (that is, an inter-distance measure) is used instead of a correlation measure between variables (as used in standard PCA), giving an alternative approach for monitoring processes that have nonlinear characteristics. Nonlinear extensions of SSA could be an alternative solution to capturing nonlinear characteristics of the process data. Specifically, the key contribution of this thesis is a set of novel multivariate statistical process monitoring approaches using nonlinear singular spectrum analysis that address limitations of conventional MSPC techniques and their extensions.

1.3 PROBLEM STATEMENT

Process monitoring can play a key role in detecting faults, identifying root causes of abnormal events and improving the performance of control systems. Better process monitoring becomes possible with more known characteristics of process signals. Therefore, the nonparametric method SSA, which decomposes data into multiple modes with a featured component, was chosen for monitoring the process signals in this study. Process monitoring based on SSA is data-adaptive where the basis function is adapted

to data that does not suffer from the limitations of other spectral techniques, where fixed-basis functions are chosen prior to data analysis. Decomposing signals on the basis of their singular spectra is equivalent to using wavelets that are constructed from the data themselves, with their shape (basis function) adapted to fit these data accurately (Yiou *et al*., 2000). However, SSA explicitly accounts for the auto-correlation in observed process data, and is particularly well suited for handling short time series as well. Although multivariate statistical process monitoring techniques based on SSA have been proven effective in process monitoring, there are still a few challenges that need attention, particularly for dynamic, nonlinear systems as are encountered in chemical engineering. A multimodal nonlinear process monitoring framework using singular spectrum analysis for signal decomposition is proposed as an alternative to existing spectral approaches for handling nonlinear data

In this study, multimodal nonlinear process monitoring techniques based on NLSSA, SSA with dissimilarity matrix (DSSA) and SSA with kernel multidimensional scaling (KDSSA), are proposed to detect process deviations at multiple modes. Thus most of the limitations of using multiscale SSA (MS-SSA) in nonlinear data can be improved by evaluating the data at the respective modes with variations in basis SSA decomposition with nonlinear projection methods, rather than linear PCA, as has been the case in previous studies. The proposed multimodal statistical monitoring strategy extends the suitability of NLSSA, DSSA and KDSSA to statistically monitor processes based on nonlinear auto-correlated measurements. In addition, the resulting nonlinear models at multiple scales are more sensitive in detecting various changes in a process. These ideas are to be illustrated by suitable simulated and industrial data. The basic idea of the methodology proposed above is that uncorrelated nonlinear principal components are analysed using conventional linear statistical analysis techniques at different modes, which is equivalent to nonlinear analysis of the original data.

The proposed data-adaptive multimodal nonlinear algorithms will perform better than the conventional MSPC methods. Monitoring the real-time operation of the petrochemical process using the proposed method will yield a minimum false alarm rate in the early detection of process faults, and this method also can grasp any slow and feeble change in the process signals at an earlier stage.

1.4 OBJECTIVES OF THE STUDY

The main objective of this study is the development of multimodal nonlinear process monitoring methods using nonlinear singular spectrum analysis for the early and reliable detection of anomalies/deviations in nonlinear chemical process systems. An analysis of the properties of the frameworks is presented. Subsequently, the proposed techniques are demonstrated using simulated and industrial data. Also, a comparative analysis between linear and nonlinear SSA is given using the Tennessee Eastman Challenge problem as benchmark data.

In brief, the objectives of this study can be listed as follows:

- Development of a nonlinear process monitoring framework using NLSSA, DSSA and KDSSA
- Development of a guidelines towards selection of the key parameters in the implementation of NLSSA, DSSA and KDSSA
- Demonstration of the algorithm using simulated and industrial data
- Comparison of process monitoring methods based on PCA and multiscale SSA with methods based on nonlinear SSA such as multimodal nonlinear SSA, multimodal DSSA and multimodal KDSSA using the Tennessee Eastman Challenge problem.

1.5 THESIS OUTLINE

The thesis is organised as follows: Chapter 2 provides a detailed illustration of methodological issues and implementation procedures of the typical data-driven process monitoring method, along with a detailed review of the advantage of the basic multivariate methods and their extensions as part of the literature review. Applications of MSPC and their limitations are also reviewed in this chapter. Chapter 3 discusses the basic procedure of singular spectrum analysis and multiscale SSA, their applications and limitations in chemical process monitoring. Alternative multimodal nonlinear monitoring strategies based on variations in the basic SSA procedure with NLPCA, MDS and kernel MDS are proposed, and the respective methodologies are demonstrated and assessed by means of three case studies: a simulated nonlinear systems, as well as data from a base metal flotation plant and the Tennessse Eastamen Challenge in Chapter 4 and Chapter 5. The conclusion of the thesis and opportunities for future research are highlighted in Chapter 6.

CHAPTER 2: MULTIVARIATE STATISTICAL PROCESS MONITORING: A LITERATURE REVIEW

In this chapter, an overview of a typical data-based process monitoring methodology and some of the state-of-the-art data-driven process monitoring methods for linear, nonlinear and auto-correlated dynamic processes are reviewed.

2.1 DATA-DRIVEN PROCESS MONITORING METHODOLOGY

Different procedures that are commonly used in data-based process monitoring schemes and their implementation through different approaches are described briefly in this section, since the development of a process monitoring model incorporates these procedures in the process monitoring task. Figure 2.1 shows an overview of the data-based process monitoring scheme, adopted from Ge *et al*. (2013) and discussed briefly in the following subsections.

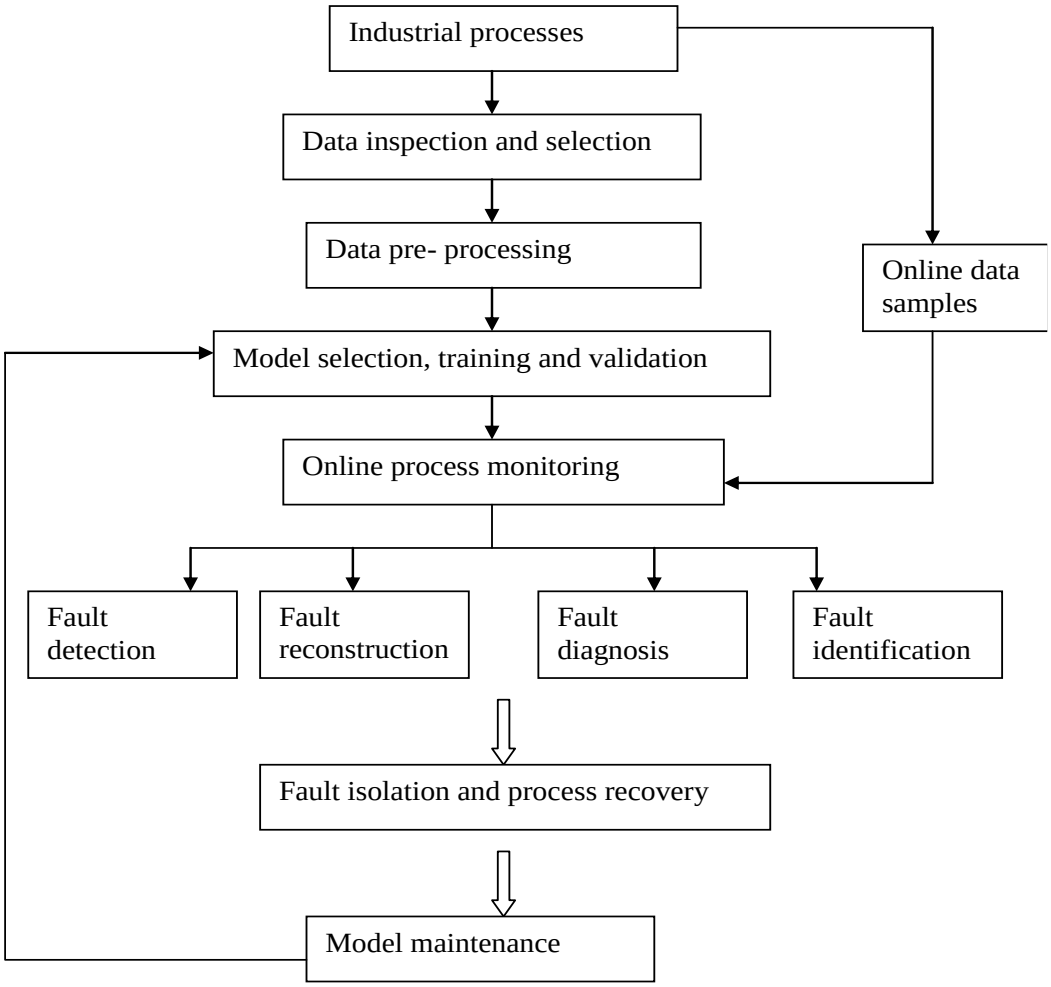


Figure 2.1: Data-based process monitoring scheme (Ge *et al*., 2013).

2.1.1 Data inspection and selection

Data inspection and selection is the first step in data-based process monitoring. In this step, the data structure is examined, different data characteristics area analysed, the operating region is identified, a suitable dataset for modelling is determined and the evaluation is performed. The success of process monitoring methods depends entirely on the proper data inspection and the selection of an appropriate dataset for modelling. Any inappropriate selection of a dataset for modelling results in a false alarm or miss-detection of process faults. Thus this step is an important one in the process monitoring scheme because the following steps are all based on it.

2.1.2 Data pre-processing

Data pre-processing transforms the original data in an appropriate manner through uniform data-scaling of the different process variables to avoid the model being inclined to any one of the process variables. It includes normalisation of data samples to zero means and unit variance in MSPC techniques. Other issues, such as gross errors in the dataset, outliers and missing data, are common in practice and can be handled through data pre-processing.

2.1.3 Model selection, training and validation

The type and complexity of the model structure is determined on the basis of the analysis of the data characteristics and the results of the evaluation. For example, if the variables are linearly correlated, a linear model should be developed, and for nonlinear correlation some nonlinear modelling approach can be employed for the best results in process monitoring. Thus the selection of a proper, optimal model is important for the better performance of the monitoring task. However, there are no specific criteria for model selection, as this is often done on the basis of experience or on the type of process. The selected model is then trained based on the process data and its performance is evaluated based on the model. For the training and evaluation steps in MSPC models, the process dataset is divided into a training dataset and a testing dataset.

2.1.4 Online process monitoring

The online monitoring task broadly involves four hierarchical tasks, namely fault detection, fault diagnosis, fault reconstruction and fault identification, based on the data-based model and appropriate statistics that are computed from the model.

In **fault detection**, the aim is to determine whether an abnormal event happened in the process or when a process or plant being monitored is out of control. The early detection of a fault condition is important in avoiding below quality product batches or system breakdown, and this can be achieved through the proper design of effective fault detection monitoring statistics with proper control limits. If the values of the monitoring statistics exceed the control limits of the corresponding statistics, a fault is detected. Once a fault condition has been positively detected, the next step is to **diagnose** or determine the root cause of the out-of-control status. The subsystems or even a sensor/actuator are mainly monitored to diagnose faults. After fault detection it is

important to know the direction, magnitude and time of occurrence of the fault, which is done by **fault reconstruction.** The normal value of the faulty data can be recovered after the fault reconstruction. This is important for fault isolation, process recovery and process control. When a fault has been detected, an understanding of the characteristics of the fault helps the operator find the proper maintenance strategy to get the process back into normal operating condition. This task is referred to as **fault identification.**

2.1.5 Fault isolation and process recovery

After detailed information on the detected faults has been obtained, the system is corrected by elimination of the fault or its cause via the fault isolation and process recovery phase. The operators isolate the fault from other parts of the process without any significant impact on process operation. The isolated fault is repaired to bring the process back to normal.

2.1.6 Model maintenance

Model maintenance is the evaluation of the feasibility and efficiency of the process monitoring model to reduce the number of false alarms due to process changes and slow drifts. Model maintenance methods that have been developed in the past include recursive/adaptive, moving-window approaches, and multimodal methods and so on.

Data-based process monitoring statistics based on multivariate methods and their applications in fault detection in industrial processes are introduced briefly in the next section.

2.2 DATA-DRIVEN PROCESS MONITORING STRATEGIES

The development of data-driven process monitoring strategies started with the advent of basic Shewhart charts, which were developed for each variable monitored on the assumption that a process is in control as long as all variables are in control. Pre-defined thresholds, called *control* limits, are used for each variable to define the boundary for in-control operations in this chart. The process is considered to be out of control if any one of the variables violates the control limits (Raich & Cinar, 1996). The other types of univariate charts, such as cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) charts, are modifications of basic Shewhart charts for the detection of small shifts in the process. The Shewhart chart, together with CUSUM and the EWMA, have been proven to be effective in monitoring single variables. Modern chemical

processes are characterised by a high prevalence of instrumentation and process automation systems that enable the tracking of a massive set of process variables. Thus, it is not uncommon to have large amounts of data collected every few seconds on such plants. The large number of measured process variables makes it extremely difficult for operators to monitor large-scale systems by simply observing the data on a univariate time series plot, such as the Shewhart, CUSUM and EWMA charts (Kourti *et al*., 1996). Moreover, as most chemical processes are dynamic, the interdependence or correlations among the variables can result in process deviations going unnoticed when traditional univariate statistical quality control approaches are used. All these limitations of univariate charts led to the development of the multivariate extensions of Shewhart control charts for the simultaneous monitoring of different variables (Hotelling, 1947).

The multivariate extensions of Shewhart control charts, such as χ^2 and Hotelling $\boldsymbol{T}^2$ control charts, are used when the process parameters of the underlying process are known and unknown respectively. For the detection of small shifts, multivariate extensions of CUSUM (MCUSUM) and EWMA (MEWMA) were developed (Lowry *et al*., 1992). When the number of variables to be monitored is large and they are highly correlated with one another, then the data is ill-conditioned in nature and the multivariate extensions of control charts fail to provide results (Das *et al*., 2012; MacGregor and Kourti, 1995). These problems can be solved with the use of data-driven multivariate monitoring methods that project high dimensional data into a lower dimensional coordinate that reveals the intrinsic patterns in the process data. As a result, multivariate projection methods such as principal component analysis (PCA) and partial least squares (PLS) have been developed and applied for more than two decades to monitor the process data (Kresta *et al*., 1991; MacGregor & Kourti, 1995). This chapter focuses on PCA and its extensions in MSPC, although the monitoring results are equally applicable to PLS. PCA describes major trends in the data by finding a linear combination of variables that maximise the explained variance in a more compact space (Das *et al*., 2012; Kresta *et al*., 1991; Wise & Gallagher, 1996; Wold *et al*., 1987; Yin *et al*., 2014).

Observed process data exhibit many properties, such as nonlinearity, serial correlation/auto-correlation and multiscale behaviour. Therefore the modelling of any system requires taking into account these characteristics in the formulation of the model.

Thus several extensions of PCA have been introduced to handle such aspects in process monitoring over the last decade. For example, nonlinear PCA and kernel PCA aim to deal with non-linearity among the process variables and can be used to capture most of the variation in the data in a reduced dimension compared to linear PCA, while dynamic PCA takes into account auto-correlation in the data and multiscale PCA exploits the multiscale characteristics of process variables. All these extensions to basic PCA might solve some of the issues that are encountered in the process, but they by no means change the nature of the implications of the non-linearity of the original multivariate data. Thus, an alternative process monitoring framework using classical multidimensional scaling (CMDS) has recently been proposed to monitor multivariate data. This method uses the inter-dissimilarity structure of the original data in the development of principal components in reduced dimensional space (Yunus, 2012).

The majority of process monitoring methods proposed in the literature employ a uniform strategy based on PCA, artificial neural networks, self-organising maps, qualitative trend analysis or signal-processing methods (Gosh, 2012). But it is difficult to develop and apply a unified method to monitor a modern chemical plant that is highly complex with a large number of process variables. As a result, each method has its own advantages and limitations and there is no method that can perform well in all chemical processes. Thus, one monitoring method that can give good monitoring performance in one particular process may not perform well in another process. As a result there is a need to develop a monitoring method that integrates with other methods to overcome their limitations.

The basic PCA process monitoring framework and its linear and nonlinear extensions in multivariate process monitoring are discussed briefly in the following subsections. As a solution to the PCA framework, an alternative multivariate technique, CMDS, and its connections to the PCA and kernel PCA are also reviewed briefly below.

2.2.1 Principal component analysis (PCA)

PCA is a basic data-driven multivariate method, originally proposed for the dimensionality reduction of large number of correlated data, and preserves the significant variability information extracted from process data. PCA has been used widely in several applications, such as image analysis, feature extraction, pattern recognition, data compression and time series prediction. PCA describes the significant

process deviations by constructing a reduced set of statistically independent variables called principal components (PCs), which are the linear combination of original variables. In PCA, the total variance of the original set of variables remains unchanged and redistributed, such that the most is in the first PC, the next largest goes to the second PC, and the least to the last PC. The advantage of using PCA in process monitoring is its ability to explain the total variation in the data as much as possible with the least number of principal components. PCA is based on an eigenvector decomposition of the covariance or correlation matrix of the process variables (Ralston *et al*., 2004; Wise & Gallagher, 1996; Yin *et al*., 2014).

For a given data matrix $\mathbf{X}$, with n observations and m process variables, the covariance matrix of $\mathbf{X}$ is approximated after normalising them to zero mean and unit variance as

$$Cov(\mathbf{X}) = n^{-1}\mathbf{X}^T\mathbf{X}\,. \qquad \textbf{(2.2.1)}$$

The eigenvalue decomposition is performed on the covariance matrix in equation (2.2.1) to obtain the significant variability information from the pre-processed data, that is,

$$n^{-1}\mathbf{X}^T\mathbf{X} = \mathbf{V}\boldsymbol{\lambda}\mathbf{V}^T \qquad \textbf{(2.2.2)}$$

where $\mathbf{V}$ is a matrix of loading or eigenvectors of size $m \times m$ and $\boldsymbol{\lambda}$ is the diagonal matrix containing the eigenvalues (λ_i), with $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_m \geq 0$ of the covariance matrix of $\mathbf{X}$. Projections of the data on the principal axes (given by columns of $\mathbf{XV}$) are called principal components or scores.

Alternatively, the scores and loadings can be computed by singular value decomposition (SVD) of the data matrix $\mathbf{X}$ by

$$\mathbf{X} = \mathbf{U}\,\boldsymbol{\Lambda}\,\mathbf{V}^T \qquad \textbf{(2.2.3)}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix containing the singular values (σ_i),$\mathbf{U}$ and $\mathbf{V}$ are unitary matrices of size $n \times n$ and $m \times m$ respectively. The loading vectors are the orthonormal column vectors in the matrix $\mathbf{V}$, and the variance of the data matrix $\mathbf{X}$ projected along the i^{th} column of $\mathbf{V}$ is equal to σ_i^2. Thus solving (2.2.3) is equivalent to solving an eigenvalue decomposition of the covariance matrix of $\mathbf{X}$ as shown in equation (2.2.2).

The PCA model of $\mathbf{X}$ can be developed by retaining k largest eigenvalues and the corresponding loading vectors to capture the variations in the data, while minimising the effect of random noise by discarding $m-k$ eigenvalues. The several of eigenvalues in a dataset with a large number of variables are zero or close to zero due to multiple measurements of the same variables, or collinearity problem due to the redundancy in the dataset. This redundancy can be eliminated from the data by choosing eigenvalues greater than a very small positive number and corresponding eigenvectors (Ku *et al*., 1995). Thus, the dimensionality reduction by retaining k eigenvalues results in a better estimate of the covariance matrix and a better feature extraction (Russell *et al*., 2000). In other words, PCA decomposes the data matrix $\mathbf{X}$ as a sum of the product of the score vectors $\mathbf{t}_i$ and PC loadings $\mathbf{v}_i$, plus a residual matrix $\mathbf{E}$:

$$\mathbf{X} = \mathbf{T}_k\mathbf{V}_k^T + \mathbf{E} = \sum_{i=1}^{k} \mathbf{t}_i\,\mathbf{v}_i^T + \mathbf{E} \qquad \textbf{(2.2.4)}$$

where $k \leq \min\{n, m\}$. The score space is represented by the approximated data matrix

$$\widehat{\mathbf{X}} = \mathbf{T}_k\mathbf{V}_k^T = \sum_{i=1}^{k} \mathbf{t}_i\,\mathbf{v}_i^T \qquad \textbf{(2.2.5)}$$

The orthogonal scores vector is the linear combination of the original data $\mathbf{X}$, defined by $\mathbf{v}_i$, that is

$$\mathbf{t}_i = \mathbf{X}\,\mathbf{v}_i \qquad \textbf{(2.2.6)}$$

which contains information on the relationship between the samples, while the orthonormal loading vectors, $\mathbf{v}_i$, contain information on the relationship between the variables that are useful for process monitoring. The eigenvalues, λ_i are the measure of the amount of variance described by the pair $\mathbf{t}_i$,$\mathbf{v}_i$, and they are arranged in descending order according to the associated λ_i. In other words, the first pair captures the largest amount of information (variation) in the data and the remaining variations in the data are captured by the subsequent pairs (Wise & Gallagher, 1996). The residual matrix,$\mathbf{E}$ contains the part of the variation in the data that is not explained by the PCA model by the first k, $\mathbf{t}_i$,$\mathbf{v}_i$ pair, where $i = 1,2,\ldots,k$. The appropriate number of eigenvalues or PCs

is selected by techniques such as percent variance, parallel analysis, scree plots of the eigen-spectrum and cross-validation (Jackson, 1991).

For the PCA applications considered later, the percent variance criteria were chosen for selecting the appropriate number of PCs to represent the measure of variation in the data. The first k eigenvalues and corresponding eigenvectors were chosen by calculating the smallest number of loading vectors needed to explain the specified minimum percentage of the total variance,$\theta_{threshold}$, typically 90% or 95%:

$$k > min\left\{d \in (1,m) / \frac{\sum_{i=1}^{d} \lambda_i}{\sum_{i=1}^{m} \lambda_i} > \theta_{threshold}\right\}. \quad \textbf{(2.2.7)}$$

The discarded eigenvalues correspond to those PCs with high-frequency variations in data, probably due to the influence of noise. A geometric representation of PCA is illustrated for a three-dimensional system in Figure 2.2, where the data are explained well by two principal components (Krishnannnair, 2010).

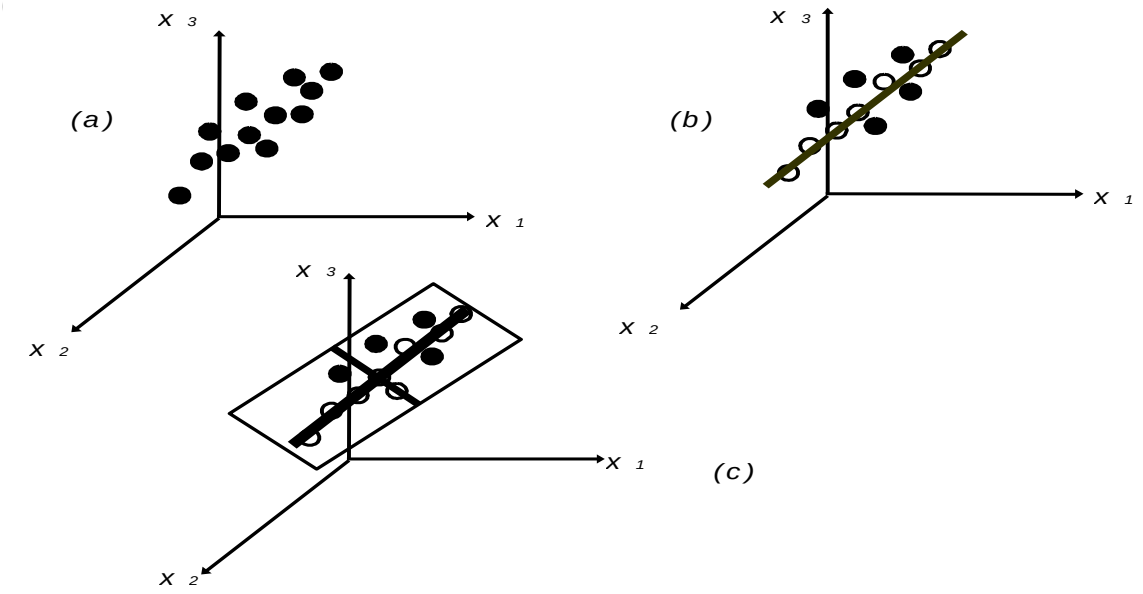


Figure 2.2: Geometric representations of the steps in the principal component analysis for a three-dimensional system showing (a) the data points in the observation space, (b) the first principal component, and (c) the plane defined by the first two principal components. This figure indicates that the derivation of principal components is based on the successive projection of lines through three-dimensional space (Krishnannair, 2010).

The process monitoring step with PCA involves the calculation of Hotelling's $\boldsymbol{T^2}$ statistic and squared prediction errors (SPE), or $\boldsymbol{Q}$ statistics. Hotelling's $\boldsymbol{T^2}$ statistic is the sum of normalised squared principal components, which explains the measure of variations within the score space and is defined as

$$\boldsymbol{T}_i^2 = \sum_{j=1}^{k} \frac{\mathbf{t}_{ij}^2}{\lambda_j} = \mathbf{x}_i \mathbf{V}_k \boldsymbol{\lambda}^{-1} \mathbf{V}_k^T \mathbf{x}_i^T \qquad \textbf{(2.2.8)}$$

where $\boldsymbol{T}_i^2$ is the $\boldsymbol{T}^2$ value for the i^{th} row of measurements, k is the number of selected PCs, $\mathbf{t}_{ij}$ is the PC corresponding to the i^{th} row and j^{th} eigenvector, $\mathbf{X}_i$ is the i^{th} observation in the data matrix, $\mathbf{X}$, $\boldsymbol{\lambda}^{-1}$ is a diagonal matrix containing the inverse eigenvalues corresponding to k retained eigenvectors, and $\mathbf{V}_k$ is the matrix of k loading vectors retained in the PCA model. Confidence limits for $\boldsymbol{T}^2$ can be calculated by using the F-distribution, as follows:

$$T^2_{k,n,\alpha} = \frac{(n-1)k}{n-k} F_{k,n-k,\alpha} \quad \textbf{(2.2.9)}$$

where $F_{k,n-k,\alpha}$ is the upper $100.\alpha$ % critical point of the F-distribution, with k and $n-k$ degrees of freedom.

If the process has shifted outside the normal operating conditions, the change in the normal condition will not be captured in the PCA model. Hence, using only the $\boldsymbol{T}^2$ chart based on the first k, PCs may not be sufficient for detecting the fault. Such process shifts may cause a change in the nature and dimensions of the relationship between the process variables, and can be detected using both the $\boldsymbol{T}^2$ statistic and the $\boldsymbol{Q}$ statistic (Kresta *et al*., 1991).

The $\boldsymbol{Q}$ statistic measures variability in the data that is not captured by the k principal components retained in the score, and is obtained from the sum of the squared errors in the residual space or the sum of variations in the residual space, which is defined as

$$\boldsymbol{Q}_i = \sum_{j=1}^{k} (x_{ij} - \hat{x}_{ij})^2 = \mathbf{e}_i \mathbf{e}_i^T = \mathbf{x}_i(\mathbf{I} - \mathbf{V}_k\mathbf{V}_k^T)\mathbf{x}_i^T \quad \textbf{(2.2.10)}$$

where $\hat{x}_{ij}$ is the predicted value of x_{ij}, $\mathbf{e}_i$, is the i^{th} row of the residual matrix $\mathbf{E}$, and $\mathbf{I}$ is the identity matrix of appropriate size. For all the given eigenvalues, λ_i, of the covariance matrix of $\mathbf{X}$, the upper confidence limit for $\boldsymbol{Q}$ can be calculated as

$$Q_\alpha = \Lambda_1 \left[1 + \frac{c_\alpha (2\Lambda_2\theta^2)^{\frac{1}{2}}}{\Lambda_1} + \frac{\Lambda_2\theta(\theta-1)}{\Lambda_1^2} \right]^{\frac{1}{\theta}} \quad \textbf{(2.2.11)}$$

where $\Lambda_i = \sum_{j=k+1}^{m} \lambda_j^i$ for $i = 1,2,3 \ldots, \theta = 1 - \frac{2\Lambda_1\Lambda_3}{(3\Lambda_2^2)}$, and c_α is the standard normal deviate corresponding to the upper (1-α) percentile.

The PCA model that represents the normal behaviour can be now used to predict the future behaviour of the process by referencing the new model based on the new observation against this controlled model. For the new set of observations, the values of the $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics are calculated by projecting them onto the plane obtained by the

k principal component loading vectors retained in the in-control PCA model. The new scores are obtained as

$$\mathbf{t}_{i,new} = \mathbf{X}_{new}\mathbf{v}_i. \quad \textbf{(2.2.12)}$$

The new residuals are calculated as follows:

$$\mathbf{E}_{new} = \mathbf{X}_{new} - \hat{\mathbf{X}}_{new} \quad \textbf{(2.2.13)}$$

where $\hat{\mathbf{X}}_{new} = \mathbf{t}_{\mathrm{k,new}}\mathbf{V}_k^T$.

For the new dataset, at a specific time, if the value of the $\boldsymbol{T}^2$ or $\boldsymbol{Q}$ statistics exceeds the control limits, the process is monitored as being out of control at that time. The large value of the $\boldsymbol{Q}$ statistics indicates the change in the correlation of current variables according to the reference PCA model of the data in the normal operating condition, and a shift in the operating condition is usually detected by the $\boldsymbol{T}^2$ chart (MacGregor & Kouriti, 1995; Qin, 2003). However, both statistics should be used simultaneously in the process monitoring scheme to detect different types of faults in the process data.

PCA has been applied successfully to several chemical processes with a large number of highly correlated variables. However, chemical process data are dynamic and contaminated by gross errors and by noise, hence the application of PCA may suppress noise to a certain level by projecting data into a lower dimension with fewer principal components due to the influence of random noise. Several extensions of MSPC approaches have been proposed to cope with the restrictions of linear PCA in process monitoring and fault diagnosis. For example, dynamic PCA has been proposed to decorrelate the auto-correlation in process data by augmenting the data matrix with time-lagged variables (Ku *et al.*, 1995; Lin *et al*., 2000; Luo *et al*., 1999). Adaptive PCA updates the model parameters continuously by exponential smoothing so as to get the model adjusted to suit new operating conditions (Wold, 1994). Multiway and multiblock PCA are suitable for batch process operations (MacGregor *et al*., 1994; Nomikosi and MacGregor, 1995; Wold et al., 1996). Moreover, multiblock PCA allows for the efficient computation of very large datasets; in model-based PCA, a first principle model is integrated with PCA (Rotem *et al*., 2000), moving PCA for monitoring changes in the direction of principal components (Kano *et al.,* 2001). The MSPC method based on the

dissimilarity of process data, called DISSIM (Kano *et al.*, 2002), is a method that incorporates external information into a PCA model based on constrained PCA (Yoon & MacGregor, 2001), and the monitoring of the performance of MSPC is further improved by the development of independent component analysis (ICA) (Kano *et al*., 2003). However, with the increased requirements of multiple products and operation conditions, the process variations in practical chemical processes are more significant, and the relationships among the process variables are nonlinear. Thus, the use of linear PCA in nonlinear problems can lead to the loss of important information in the data if the minor principal components are discarded. This is because the minor components in nonlinear data always carry important information (Xu *et al*., 1992). On the other hand, PCA might contain too many principal components to solve an application if the minor components are not discarded. In general, a large number of principal components (PCs) are necessary for nonlinear systems to capture the manifold structure(s) in the data (Zhang *et al*., 1997). To address this challenge, nonlinear extensions to PCA have been proposed that are well suited when nonlinear correlations are present in the observed data (Dong & McAvoy, 1996).

The objective of nonlinear PCA (NLPCA) is to capture both linear and nonlinear relationships between process variables. This objective is achieved by projecting the process variables onto curves or surfaces instead of lines or planes (Harkat *et al.*, 2007). The concepts of linear principal component and nonlinear principal component are illustrated in Figure 2.3(a) and (b) respectively. In Figure 2.3(a), the one-dimensional subspaces can be approximated by two principal components, whereas in Figure 2.3(b) they can be approximated using a continuous curve or one nonlinear component. Moreover, Figure 2.3(a) shows that the linear principal component minimises the sum of the squared orthogonal deviations using a straight line, while Figure 2.3(b) shows that the nonlinear principal component minimises the sum of squared orthogonal deviations using a smooth curve.

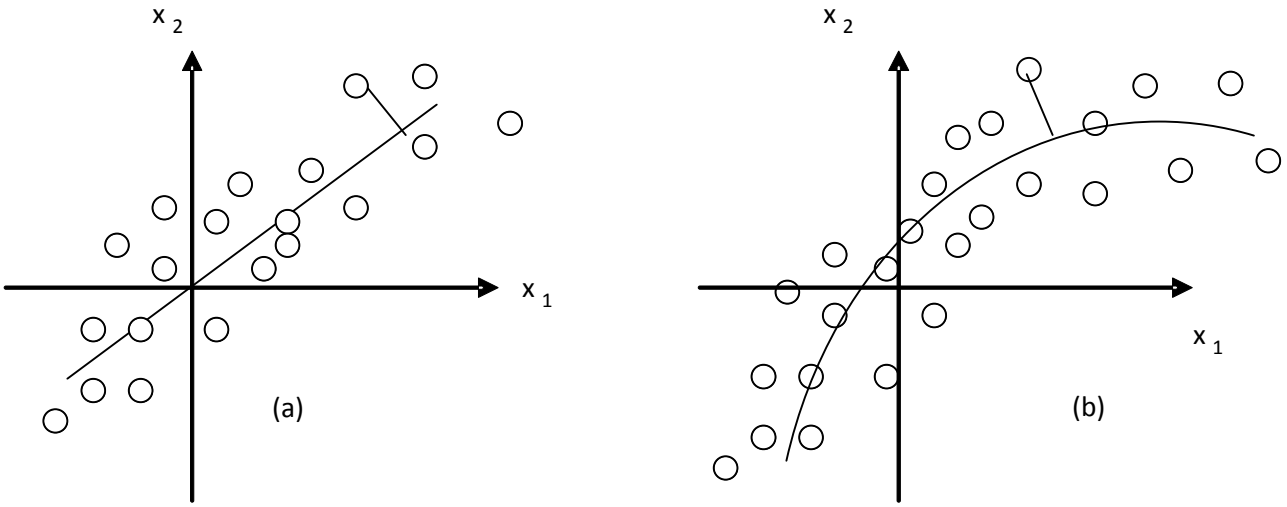


Figure 2.3: Illustration of the concept of (a) linear PCA and (b) nonlinear PCA. A straight line is used in PCA to minimize the sum of squared orthogonal deviations whlile NLPCA used a smooth curve to minimize the sum of the orthogonal deviations.

To overcome the limitations of linear MSPC techniques, nonlinear extensions of PCA, such as nonlinear PCA based on auto-associative neural networks and principal curves, have been proposed for process monitoring (Dong & McAvoy, 1996; Kramer, 1991). In both these methods, a nonlinear optimisation problem is solved that is costly and not guaranteed to find the global optimal solution. Kernel principal component analysis (KPCA), a nonlinear dimension-reduction technique, was proposed as an alternative that avoids the computational complexity of the above nonlinear PCA frameworks by solving a linear algebraic formulation with a global optimum solution (Choi & Lee, 2004; Choi *et al.*, 2005; Lee *et al*., 2004).

In nonlinear processes, fault detection is more complicated than in linear processes, and fault may get smeared in the process. There is no one such unified method that can detect the faults in all industrial processes; this is because of the differences in the nonlinear relationship among the process variables. One method may work well in one process but may not function well in another process (Ge *et al*., 2013). Although many different methods to extract nonlinear principal components are available in the literature, the nonlinear PCA based on auto-associative neural network and kernel PCA was used in this study to extract nonlinear components from the data. Detailed reviews of these methods are provided in following sections (Choi *et al*., 2005; Kramer, 1991; Lee *et al*., 2004).

2.2.2 Nonlinear PCA (NLPCA)

For a data matrix, $\mathbf{X} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n]^T \in \mathbf{R}^{n \times m}$, with n observations and m variables, as discussed earlier, PCA reconstructs the original data matrix, $\widehat{\mathbf{X}}$, by finding a projection axis $\mathbf{w} \in \mathbf{R}^{m \times 1}$ that represents the direction with maximum variability to represent the original data (Jackson, 1991). That is,

$$\max_{\mathrm{w}} var(\mathbf{X}\mathbf{w}), subject\ to\ \|\mathbf{w}\|^2 = 1\,. \qquad \textbf{(2.2.14)}$$

The above objective function is equivalent to the minimum reconstruction error criterion (Sanger, 1989; Zhao & Xu, 2005).

$$J_1 = \min_{\mathrm{w}} \frac{1}{n} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{x}_\mathrm{i}\mathbf{w}_i\mathbf{w}_i^T\|^2. \qquad \textbf{(2.2.15)}$$

After obtaining the first principal component, $\mathbf{X}\mathbf{w}_1$, the original data $\mathbf{X}$ is deflated as $\mathbf{X}_1 = \mathbf{X} - \mathbf{X}\mathbf{w}_1\mathbf{w}_1^T$. The remaining PCs are then retrieved by using $\mathbf{X}_1$ in the same manner and the process is repeated until a convergence index is satisfied (Zhao & Xu, 2005).

The above standard approach in PCA has been extended to develop NLPCA to find both the linear and nonlinear relationships between the variables by projecting the process variables onto curves or surfaces instead of lines or planes by minimising the mean square error, $E\{\|\mathbf{X} - \widehat{\mathbf{X}}\|\}$ (Shao *et al*., 1999), as illustrated in Figure 2.3(b).

The NLPCA models can be represented by mapping and de-mapping sub-models. The nonlinear principal component, $\mathbf{T}$, is obtained from the mapping model, while the de-mapping model gives the estimation of the data matrix, $\mathbf{X}$. The nonlinear mapping is expressed as

$$\mathbf{t} = \boldsymbol{G}(\mathbf{x}) \qquad \textbf{(2.2.16)}$$

where $\mathbf{x} \in \mathbf{X}$ and $\mathbf{t} \in \mathbf{T}$, and $\boldsymbol{G}$ is the nonlinear mapping function. The de-mapping model estimates $\hat{\mathbf{x}}$ of $\mathbf{x}$ from the nonlinear principal component, $\mathbf{t}$, by inverse transformation and is given as

$$\hat{\mathbf{x}} = \boldsymbol{F}(\mathbf{t}) \qquad \textbf{(2.2.17)}$$

where $\boldsymbol{F}$ is the nonlinear de-mapping function, which is equivalent to the projection matrix $\mathbf{V}$ in linear PCA (Shao *et al*., 1999).

Therefore, the original data, $\mathbf{X}$, can be represented by k nonlinear principal component as

$$\mathbf{X} = \widehat{\mathbf{X}} + \mathbf{E} = \boldsymbol{F}(\mathbf{T}) + \mathbf{E} \tag{2.2.18}$$

where $\mathbf{T} = [\mathbf{T}_1, \dots, \mathbf{T}_k]$ is the matrix of nonlinear principal components $\mathbf{T} = \boldsymbol{G}(\mathbf{X})$, and $\mathbf{E}$ is the residual matrix.

The nonlinear projection functions $\boldsymbol{G}$ and $\boldsymbol{F}$ are identified by minimising the objective function, which is the sum of squared orthogonal deviations, i.e.

$$\min \sum_{i=1}^{n} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2 = \min \sum_{i=1}^{n} \|\mathbf{x}_i - \boldsymbol{F}(\boldsymbol{G}(\mathbf{x}_i))\|^2 \tag{2.2.19}$$

where $\mathbf{x}_i$ is the i^{th} row of $\mathbf{X}$ and $\hat{\mathbf{x}}_i$ is its estimation by the five-layer neural network NLPCA model (Harkat *et al.*, 2007; Kramer, 1991).

The NLPCA proposed by Kramer (1991) is based on multilayer perception (MLP) with an auto-associative neural network. The standard auto-associative neural network consists of three hidden layers of neurons between the input and output layers of variables, which gives five layers altogether, with a single bottleneck layer that executes an identity projection, as shown in Figure 2.4 (Hsieh, 2007).

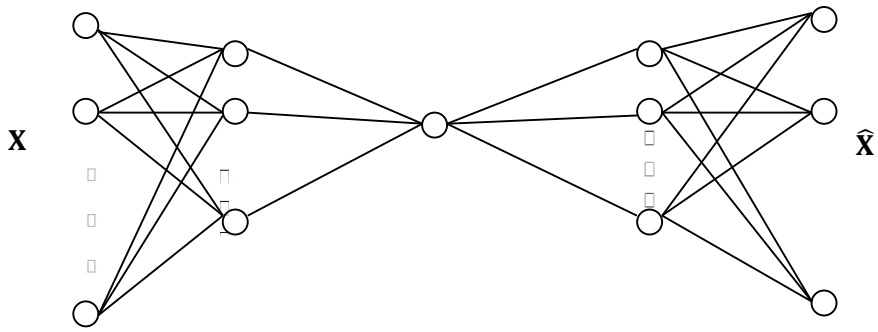


Figure 2.4: Schematic diagram of an auto-associative neural network (Hsieh, 2007). There are three layers of hidden neurons between the input layer $\mathbf{x}$ and the output layer $\hat{\mathbf{x}}$.

The input column vector $\mathbf{x}$ of length m is mapped to the first hidden layer $\boldsymbol{h}^{(\mathbf{x})}$ of network of length r by a transfer function, $\boldsymbol{G}_1$, with elements

$$\boldsymbol{h}_j^{(\mathbf{x})} = \boldsymbol{G}_1\left(\left(\mathbf{V}^{(\mathbf{x})}\mathbf{x} + \boldsymbol{b}^{(\mathbf{x})}\right)\right) = \boldsymbol{G}_1\left(\sum_{i=1}^{m} \mathbf{v}_{ij}^{\mathbf{x}}\,\mathbf{x}_i + \boldsymbol{b}_j^{(\mathbf{x})}\right) \qquad \textbf{(2.2.20)}$$

where $\mathbf{V}^{(\mathbf{x})}$ is an $r \times m$ weight matrix, $\boldsymbol{b}^{(\mathbf{x})}$, column vector, containing a bias parameter of length r, and $j = 1, 2, \ldots, r$.

The nonlinear principal component $\mathbf{t}$ is computed as

$$\mathbf{t} = \boldsymbol{w}^{(\mathbf{x})}\boldsymbol{h}^{(\mathbf{x})} + \overline{\boldsymbol{b}}^{(\mathbf{x})} = \sum_{j=1}^{r} \boldsymbol{w}_j^{(\mathbf{x})}\boldsymbol{h}_j^{(\mathbf{x})} + \overline{\boldsymbol{b}}^{(\mathbf{x})}\,. \qquad \textbf{(2.2.21)}$$

Thus the mapping function $\boldsymbol{G}$ is given as

$$\boldsymbol{G}(\mathbf{x}) = \sum_{j=1}^{r} \boldsymbol{w}_j^{(\mathbf{x})}\boldsymbol{h}_j^{(\mathbf{x})} + \overline{\boldsymbol{b}}^{(\mathbf{x})}. \qquad \textbf{(2.2.22)}$$

In the next step, the nonlinear principal component, $\mathbf{t}$, is mapped to the final hidden layer $\boldsymbol{h}^{(\mathbf{t})}$ of length r by a transfer function $\boldsymbol{F}_1$ with elements

$$\boldsymbol{h}_j^{(\mathbf{t})} = \boldsymbol{F}_1\left(\left(\boldsymbol{w}^{(\mathbf{t})}\mathbf{t} + \boldsymbol{b}^{(\mathbf{t})}\right)_j\right) = \boldsymbol{F}_1\left(\boldsymbol{w}_j^{(\mathbf{t})}\mathbf{t} + \boldsymbol{b}_j^{(\mathbf{t})}\right). \qquad \textbf{(2.2.23)}$$

Hence the demapping function $\boldsymbol{F}$ is given by

$$\boldsymbol{F}(\mathbf{t}) = \sum_{j=1}^{r} \mathbf{v}_{ji}^{(\mathbf{t})}\boldsymbol{h}_j^{(\mathbf{t})} + \overline{\boldsymbol{b}}_i^{(\mathbf{t})}. \qquad \textbf{(2.2.24)}$$

The nonlinear transfer functions, $\boldsymbol{G}_1$ and $\boldsymbol{F}_1$, are generally the hyperbolic tangent or the sigmoidal functions.

The original data, $\mathbf{x}$, is approximated by the output of the network, $\hat{\mathbf{x}}$, as

$$\hat{\mathbf{x}} = \boldsymbol{F}(\mathbf{t}) = \boldsymbol{F}(\boldsymbol{G}(\mathbf{x})) \ . \qquad \textbf{(2.2.25)}$$

By finding the optimal values of $\mathbf{V}^{(\mathbf{x})}$, $\boldsymbol{b}^{(\mathbf{x})}$, $\boldsymbol{w}^{(\mathbf{x})}, \bar{\boldsymbol{b}}^{(\mathbf{x})}, \boldsymbol{w}^{(\mathbf{t})}$, $\boldsymbol{b}^{(\mathbf{t})}$, $\mathbf{V}^{(\mathbf{t})}$ and $\bar{\boldsymbol{b}}^{(\mathbf{t})}$, the cost function, $\boldsymbol{E} = \|\mathbf{x} - \hat{\mathbf{x}}\|^2$, can be minimised. Thus the mean square error (*m.s.e*) between the neural network output $\hat{\mathbf{x}}$ and the input $\mathbf{x}$ is minimised (Harkat *et al.*, 2007). The efficient computation of latent variables in NLPCA depends on the nonlinear optimisation of the above parameters, which is complex to compute, time consuming and costly. However, the application of NLPCA is computationally demanding, since the development of the NLPCA model requires a large amount of computational time. Complications also arise in the decision on the number of mapping and de-mapping layers in the auto-associative architect for the development of the optimum model. Nevertheless, several nonlinear principal component analysis methods have been proposed in the literature to improve the feature extraction when the variables are nonlinearly correlated.

Hastie and Stuetzle (1989) proposed a principal curve methodology that provides a nonlinear summary of an m-dimensional dataset. However, this non-parametric technique cannot be used for the continuous mapping of new data. Kramer (1991) proposed nonlinear principal component analysis (NLPCA) to capture nonlinear relationships among variables. The main difference between PCA and NLPCA is that the latter uses nonlinear mapping between the original and the reduced dimension space. Compared to linear PCA, NLPCA can explain more variance in smaller dimensions (Dong & McAvoy, 1996; Kramer, 1991; Tan and Mavrovouniotis, 1995). NLPCA can be performed by a variety of methods. For example, Kramer (1991) and Hsieh (2004) used the auto-associative neural network model (ANN) to perform a nonlinear data reduction similar to PCA. In the proposed methodology, a neural network containing a bottleneck layer is used to perform the identity mapping. The network architecture makes use of three hidden layers, which are necessary to achieve the general nonlinear fitting property (Kramer, 1991). Dong and MacAvoy (1996) proposed NLPCA based on the principal curves and neural network methods to generate nonlinear principal scores. Tan

and Mavrovouniotis (1995) developed a nonlinear data-reduction technique based on the optimisation of the neural network inputs, where each input pattern is fixed but adjusted along with internal network parameters to reproduce a corresponding output pattern based on the steepest gradient descent network optimisation rule. A nonlinearity measure for the principal component is proposed to determine whether the process monitoring task used should be linear or nonlinear (Kruger *et al*., 2005). A hybrid neural network model that was developed for rule generation for processes has been also used for process monitoring (Tan *et al.*, 2007). For the online monitoring of process mean and variance shifts, an ensemble learning method has been combined with a neural network (Wu & Yu, 2010). A hierarchical neural network based on the fuzzy clustering method has been proposed for fault diagnosis in the Tennessee-Eastman benchmark process (Eslamloueyan, 2011). Recently, several other nonlinear process monitoring methods based on principal curves and neural networks have been used in systems monitoring (Antory *et al*., 2008; Jia *et al*., 2001; Karpenko *et al*., 2003; Kim *et al*., 2009; Silva, 2010).

Kernel PCA (KPCA), which is the reformulation of conventional PCA in a high-dimensional space that is constructed using kernel function, has been used in process monitoring to avoid the complexity of nonlinear optimisation in nonlinear algorithms with neural networks (Lee *et al.*, 2004). The basic concept of KPCA is that, without prior knowledge of the nonlinear function, a kernel function is used to map the measurements in the original space to a feature space or the kernel space. Then PCA is performed in the feature space to extract the nonlinear components in the data without the need to specify the number of components to be extracted prior to modelling. Thus, the application of PCA in the feature space provides kernel PCA with the property of constructing nonlinear mappings. In KPCA, the principal eigenvectors are computed using a kernel matrix, rather than the covariance matrix of the data in the feature space (Van der Maaten *et al*., 2008).

2.2.3 Kernel PCA (KPCA)

KPCA is a nonlinear PCA algorithm that maps the original input data into a high-dimensional feature space, $\mathcal{F}$ where a linear PCA model is developed (Lee *et al*., 2004; Schölkopf *et al*., 1998). The KPCA algorithm shown in this section is according to Schölkopf *et al*., (1998) and Lee *et al*. (2004). As indicated before, the KPCA algorithm involves steps such as mapping the data from the input space to a feature space and the

application of PCA in the feature space. For a given set of n training samples, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbf{R}^m$, consider a nonlinear mapping of $\mathbf{x}_j$ $(j = 1,2, \dots, n)$ into a feature space $\mathcal{F}$ by $\boldsymbol{\phi}: \mathbf{x}_j \in \mathbf{R}^m \rightarrow \boldsymbol{\phi}(\mathbf{x}_j) \in \mathcal{F}$. Then the sample covariance matrix in the feature space is

$$\mathbf{C}^f = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{\phi}(\mathbf{x}_j)\boldsymbol{\phi}(\mathbf{x}_j)^T \qquad \textbf{(2.2.26)}$$

where $\mathbf{x}_j \in \mathbf{X}$ and it is assumed that $\boldsymbol{\phi}(\mathbf{x}_j)$ is the centred nonlinear mapping of the input vector, $\mathbf{x}_j$,that is $\sum_{i=1}^{n} \boldsymbol{\phi}(\mathbf{x}_i) = 0$.

The eigenvalue problem in the feature space can be solved by diagonalising the covariance matrix, $\mathbf{C}^f$,

$$\lambda \mathbf{v} = \mathbf{C}^f \mathbf{v} \qquad \textbf{(2.2.27)}$$

where λ is the eigenvalue, $\lambda \geq 0$ and $\mathbf{v}$ is the eigenvector in $\mathcal{F} \setminus \{0\}$ .The first PC in $\mathcal{F}$ is in the direction of $\mathbf{v}$ with the largest eigenvalue, and the last PC is in the direction of the smallest eigenvalue.

Multiplying $\boldsymbol{\phi}(\mathbf{x}_k)$ on both sides of equation (2.2.27), we get:

$$\lambda \langle \boldsymbol{\phi}(\mathbf{x}_k), \mathbf{v} \rangle = \langle \boldsymbol{\phi}(\mathbf{x}_k), \mathbf{C}^f \mathbf{v} \rangle \qquad \textbf{(2.2.28)}$$

where $\langle \boldsymbol{\phi}(\mathbf{x}_j), \mathbf{v} \rangle$ is the dot product between $\boldsymbol{\phi}(\mathbf{x}_j)$ and $\mathbf{v}$, $k = 1, 2, \dots, n$ .This shows that $\mathbf{v}$ is spanned by $\boldsymbol{\phi}(\mathbf{x}_1), \boldsymbol{\phi}(\mathbf{x}_2), \dots, \boldsymbol{\phi}(\mathbf{x}_n)$, and there is a coefficient α_i $(i = 1,2, \dots, n)$, such that

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \boldsymbol{\phi}(\mathbf{x}_i). \qquad \textbf{(2.2.29)}$$

Then, by combining equations (2.2.26) and (2.2.28) and substituting in equation (2.2.29), we get

$$\lambda \sum_{i=1}^{n} \alpha_i \langle \boldsymbol{\phi}(\mathbf{x}_k), \boldsymbol{\phi}(\mathbf{x}_i) \rangle = \frac{1}{n}\sum_{i=1}^{n} \alpha_i \langle \boldsymbol{\phi}(\mathbf{x}_k), \sum_{j=1}^{n} \boldsymbol{\phi}(\mathbf{x}_j) \rangle \langle \boldsymbol{\phi}(\mathbf{x}_j), \boldsymbol{\phi}(\mathbf{x}_i) \rangle \qquad \textbf{(2.2.30)}$$

where $k = 1, 2, \dots, n$. The eigenvalue problem in equation (2.2.30) only requires the dot products of the mapped vectors in the feature space, and that can be computed using a kernel matrix. Thus, the computational difficulties in the higher dimensional feature space can be simplified into finding the dot products of vectors in the feature space by using a kernel function (Jemwa & Aldrich, 2006).

The kernel function implicitly determines the nonlinear mapping and the feature space. The kernel matrix, $\boldsymbol{K}(n \times n)$, is defined as $\boldsymbol{K}_{ij} = \boldsymbol{K}(\mathbf{x}_i, \mathbf{x}_j) = \langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j) \rangle$, and then the equation (2.2.30) can be written as

$$\lambda n \boldsymbol{K}\mathbf{u} = \boldsymbol{K}^2\mathbf{u} \tag{2.2.31}$$

where $\mathbf{u} = [\mathbf{u_1}, \dots, \mathbf{u_n}]$. As discussed by Schölkopf *et al*. (1998), mean-centring in high-dimensional space is required for the prior application of KPCA, which is done by substituting the kernel matrix, $\boldsymbol{K}$, with the following:

$$\widetilde{\boldsymbol{K}} = \boldsymbol{K} - \mathbf{1}_n\boldsymbol{K} - \boldsymbol{K}\mathbf{1}_n + \mathbf{1}_n\boldsymbol{K}\mathbf{1}_n \tag{2.2.32}$$

where

$$\mathbf{1}_n = \frac{1}{n}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \in \mathbf{R}^{n \times n}. \tag{2.2.33}$$

The solution of equation (2.2.31) can be obtained by solving the eigenvalue problem of the centered kernel,

$$n\lambda\mathbf{u} = \widetilde{\boldsymbol{K}}\mathbf{u} \tag{2.2.34}$$

for all nonzero eigenvalues, λ. Thus the application of PCA in $\mathcal{F}$ gives eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$, with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$. The dimensionality of the data is reduced by retaining the first k eigenvectors, and the solution is further normalised by imposing $\lambda_k(\mathbf{u}_k, \mathbf{u}_k) = 1$ in $\mathcal{F}$ by using $\mathbf{v}_p = \sum_{i=1}^{n} \mathbf{u}_i^p \boldsymbol{\phi}(\mathbf{x}_i)$, where $\langle \mathbf{v}_p, \mathbf{v}_p \rangle = 1$ for all $p = 1,2, \dots, k$. The principal component, $\mathbf{t}_{k,new}$, of a test vector $\mathbf{X}_{new}$ in $\mathcal{F}$ is extracted by projecting $\boldsymbol{\phi}(\mathbf{x}_{new})$ onto eigenvectors $\mathbf{v}_p$ in $\mathcal{F}$, where $p = 1,2, \dots, k$. That is

$$\mathbf{t}_{p,new} = \langle \mathbf{v}_p, \boldsymbol{\phi}(\mathbf{x}_{new}) \rangle = \sum_{i=1}^{n} \mathbf{u}_i^p \langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_{new}) \rangle. \quad \textbf{(2.2.35)}$$

In short, using integral operator kernel functions as a nonlinear map instead of the basic dot product, one can perform PCA in a high-dimensional space, $\mathcal{F}$, which is nonlinearly related to the input space, as shown in Figure 2.5 (Schölkopf *et al*., 1998). Similar to linear PCA, the principal component in KPCA is uncorrelated and orthogonal, and the first PC carries more variance than the other components. Mostly in KPCA the number of generated PCs are equal to the number of sample points in the data. Kernel functions that have been used successfully in the machine learning literature include the

Linear kernel:

$$\boldsymbol{k}(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle \quad \textbf{(2.2.36)}$$

Polynomial kernel:

$$\boldsymbol{k}(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle^d \quad \textbf{(2.2.37)}$$

Sigmoid kernel:

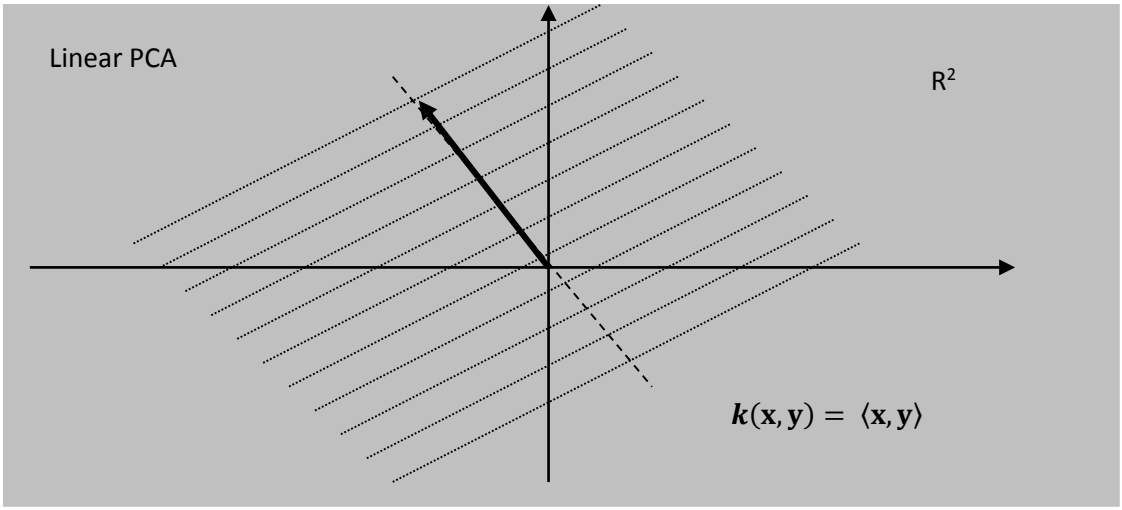
$$\boldsymbol{k}(\mathbf{x}, \mathbf{y}) = \tanh(\beta_0 \langle \mathbf{x}, \mathbf{y} \rangle + \beta_1) \quad \textbf{(2.2.38)}$$

and *Gaussian kernel:*

$$\boldsymbol{k}(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{c}\right) \quad \textbf{(2.2.39)}$$

where d, β_0, β_1 and c are specified *a priori* before using the above kernel functions and give similar results if appropriate parameters are chosen. The above kernel functions give a low-dimensional KPCA subspace that represents the distribution of the mapping of the nonlinear mapping of the training vectors in the feature space $\mathcal{F}$(Lee *et al*., 2004). The Gaussian kernel is used in this thesis to capture the nonlinear behaviour of the process data, since this distance-based kernel has the flexibility of generalising the data without any prior assumptions regarding relationships between the variables and is widely used in the process monitoring literature. The main advantage of KPCA compared to NLPCA is that it does not need nonlinear optimisation and requires only the solution of an eigenvalue problem. KPCA can handle a wide range of nonlinearities due

to its ability to use different kernels. In addition, KPCA has performed better than linear PCA methods in feature extraction and classification in nonlinear systems due to the fact that it specifies the number of components to be extracted prior to the modelling (Lee *et al*., 2004).



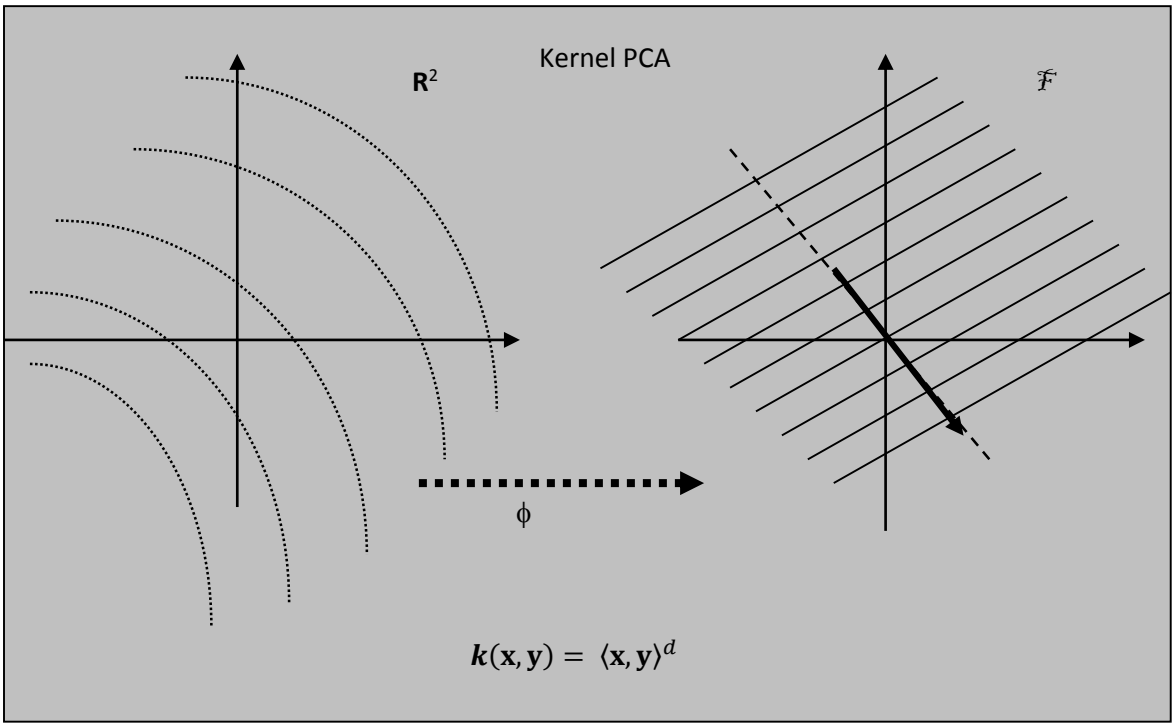


Figure 2.5: The basic idea of KPCA (Schölkopf *et al*., 1998). Nonlinear kernel functions are used to perform PCA in the feature space $\mathbb{F}$ which is nonlinearly related to the input space $\mathbf{R}^2$.

To monitor multivariate data with nonlinear characteristics, several variants of KPCA have been proposed in the literature. Lee *et al*. (2004; 2005) proposed a nonlinear

process monitoring strategy based on KPCA, with two monitoring statistics for fault detection and the identification of an industrial chemical process with different fault conditions. The technique of local approach has been combined with KPCA for an improved KPCA process monitoring scheme to handle the nonlinear and non-Gaussian characteristics of the data simultaneously (Ge *et al*., 2009). Several other methods based on KPCA involve the reconstruction-based contribution approach of Alcala and Qin (2010), the adaptive KPCA for small disturbances of nonlinear process of Cheng *et al.*, (2010), and dynamic kernel PCA for nonlinear dynamic process monitoring of Choi and Lee (2004). In some applications, KPCA has been combined with PLS, independent component analysis, support vector machines, recursive-weighted PCA and wavelets for improved process monitoring and fault diagnosis (Chakour *et al*., 2015; Choi *et al*., 2008; Lee *et al*., 2007; Zhang, 2008; Zhang & Ma, 2011; Zhang *et al.*, 2010).

Even though the computation of the KPCA model is less complicated, the implementation process involves high-dimensional models, thus the original problem of high dimensionality cannot be solved effectively. Nevertheless, nonlinear PCA models are effective in detecting faults, but the fault identification process may give false conclusions due to the inherent behaviour of principal component models (Yunus & Zhang, 2010). As an alternative to projection to latent structure techniques based on PCA, a MSPC framework that uses classical multidimensional scaling (CMDS), one of the main techniques in multidimensional scaling (MDS) (Cox & Cox, 1994), has recently been proposed in which a summary of the multivariate data is obtained in a relatively smaller dimension (Yunus &Zhang, 2010; Yunus & Zhang, 2014). In CMDS, the dissimilarity of inter-object distance is the basic measure for multivariate scores development in the data. This can also handle nonlinear correlations in the data, like other nonlinear methods in process monitoring (Yunus & Zhang, 2010). In MDS, the process signals are represented as points in a co-ordinate system, in such a way that the distance between points conveys something about the dissimilarities among the signals. In general, nonlinear techniques based on MDS can represent the high-dimensional data to a lower dimension while retaining the pairwise distance between the data points as much as possible. In particular, MDS can be considered as a method that provides the graphical representation of point configuration, which shows the predefined set of inter-distances or dissimilarities between points of reference from particular multivariate data. Moreover, Takane (2003) states that points are arranged in such a

way that their distances correspond to the correlation between the variables, and if the points have high similarity they are located closely together, otherwise they are far apart. CMDS is strongly related to PCA by the procedure for finding eigenvectors of the dissimilarity matrix that is obtained from the basis of the dot product of the original data, as done in the PCA procedure. When Euclidean distance is used as the dissimilarity measure in CMDS, then both PCA and CMDS share the same observation-sample score configuration (Cox & Cox, 1994; Yunus & Zhang, 2010). Thus, the issue of non-linearity is naturally incorporated in CMDS scores by using the dissimilarity measure, and better fault detection is possible in a reduced dimension because CMDS uses dissimilarity scale transformation instead of variance transformation in PCA. That is, in CMDS, the amount of variances will be calculated based on the multivariate score behaviours, instead of adopted from the original data as in PCA. This motivated the development of an alternative multivariate statistical nonlinear process monitoring framework based on classical multidimensional scaling to detect faults in the process data (Yunus & Zhang, 2010; Yunus & Zhang, 2014).

The first step in CMDS is the computation of a dissimilarity matrix, a matrix with values such as dissimilarity coefficients that contains information about the degree of resemblance between variables in the data. CMDS is related to PCA in the sense that the procedure for obtaining the eigenvectors of dissimilarity matrix in this study is constructed based on the Euclidean distance between the points in the dataset. In the second phase, the multidimensional dissimilarity matrix is converted into a manageable, lower dimensional configuration of points specified by coordinates through a multidimensional scaling process, which is much easier to interpret and portray (Prentice, 1980). In this study, a multivariate statistical process monitoring framework based on CMDS is used to project the multivariate scores by using the dissimilarity values of the variables and, hence, the complex computation of nonlinear principal components could be avoided.

2.2.4 Classical multidimensional scaling (MDS)

Multidimensional scaling (MDS) is a widely used multivariate statistical method for dimensionality reduction and the visualisation of similarities and dissimilarities in multivariate time series analysis (Cox, 2001). MDS reduces the dimension of a multidimensional dataset by representing it as points in a graphical coordinate system, in

such a way that the inter-distance or dissimilarity relationships among the data points are retained in a low-dimensional space (typically two- or three-dimensional for visualisation purposes). The distances between the points correspond to the correlation between the variables – points with high similarity are positioned closely together, otherwise they are further apart (Takane, 2003). Therefore, the goal of MDS is to find a configuration to measure how well the projected multivariate scores match as accurately as possible according to the predefined dissimilarity scales that corresponds with the correlation between the measurements (Kruskal & Wish, 1978). The different types of matching give rise to the different techniques of MDS, of which classical multidimensional scaling (CMDS) has recently been used to build a MSPC framework for industrial process monitoring (Yunus & Zhang, 2010).

The first step in classical MDS is the construction of the dissimilarity matrix,$\mathbf{D}$, based on the Euclidean distance between the points in the dataset (Cox & Cox, 1994). The Euclidean distance between variables i and j in a data matrix $\mathbf{X}$ with n points and m variables is defined as

$$d_{ij}^2 = \left\{\sum_{r=1}^{m}(x_{ir} - x_{jr})^2\right\} = \left(x_i - x_j\right)^T\left(x_i - x_j\right) = x_i^T x_i + x_j^T x_j - 2x_i^T x_j. \qquad \textbf{(2.2.40)}$$

In the second phase, the multidimensional dissimilarity matrix is converted into a lower dimensional configuration of points by computing the principal components from the inter-distance scale structure. Cox and Cox (1994) started the procedures for obtaining the principal component loadings from the inter-distance scale structure by defining the scalar product matrix, $\mathbf{B}$ (or the dissimilarity matrix of d_{ij}^2), that is $\mathbf{B} = \mathbf{X}\mathbf{X}^T$, where $\mathbf{X}$ is a $n \times m$ mean-centred data matrix. Through eigenvalue decomposition on $\mathbf{B}$, the eigenvalues, λ_i, and corresponding eigenvectors $\mathbf{v}_i$ of $\mathbf{B}$, can be computed, where $i = 1,2,\ldots,m$. Thus

$$\mathbf{B}\mathbf{v}_i = \lambda_i \mathbf{v}_i. \qquad \textbf{(2.2.41)}$$

From equation (2.2.41), the principal component loading factors can be computed and the computation is shown in the following equations:

$$\mathbf{X}^T\mathbf{B}\mathbf{v}_i = \lambda_i \mathbf{X}^T\mathbf{v}_i \qquad \textbf{(2.2.42)}$$

$$\mathbf{C}\mathbf{a}_i^* = \lambda_i \mathbf{a}_i^* \tag{2.2.43}$$

where $\mathbf{C} = \mathbf{X}^T\mathbf{X}$ and $\mathbf{a}_i^* = \mathbf{X}^T\mathbf{v}_i$. The orthogonal loading factors (principal components), $\mathbf{a}_i$, similar to PCA, are obtained by normalising eigenvectors $\mathbf{a}_i^*$ by the corresponding eigenvalues, λ_i,

$$\mathbf{a}_i = \frac{1}{\sqrt{\lambda_i}} \mathbf{X}^T\mathbf{v}_i \tag{2.2.44}$$

and therefore

$$\mathbf{a}_i^T\mathbf{a}_i = 1. \tag{2.2.45}$$

Thus, through the above equations, principal component analysis can be considered as a specific variant of CMDS, where the former uses the correlation measure and the latter uses the dissimilarity measure.

The matrix $\mathbf{B}$ can be computed from any means of inter-distance measures between the variables. Thus any type of inter-distance scales can be transformed into $\mathbf{B}$, based on the relationship between the scalar product, $\mathbf{B}$, and the Euclidean distance,$\mathbf{D}$, with size $n \times n$, as shown in equation (2.2.46) (Borg & Groenen, 1997; Cox & Cox, 1994; Yunus & Zhang, 2014)

$$b_{ij} = -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 - \sum_{j=1}^{n} d_{ij}^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 \right) \tag{2.2.46}$$

where, b_{ij} is the scalar product between variables i and j.

However, the projection of multivariate scores is always rearranged in the form of Euclidean space (Cartesian coordinates), and the computation of $\mathbf{B}$ using equation (2.2.46) can be regarded as an equally good or better transformation of any inter-distance scales into Euclidean distance based scales (Yunus & Zhang, 2014).

The key advantage of CMDS is its efficient computation of relatively low-dimensional data with a better visualisation of the hidden, nonlinear structure of the high-dimensional data. In CMDS, the dissimilarity (or similarity) measure is usually Euclidean distance (or the positive definite metric). Since the inner product function that induces the similarity is

positive definite, the theory of reproducing kernels can be applied, allowing for more flexibility in the choice of the dissimilarity measure (Zhang & Jordan, 2009). More specifically, in CMDS the input points are embedded into an idealised Euclidean space, on which the Euclidean distance reflects the dissimilarity/similarity between the points in the input space. This embedded Euclidean space and points are also considered as feature space and feature vectors. The points in the input space can be embedded into the feature space with linear or nonlinear mapping (kernel functions). Thus CMDS is a nonlinear dimensionality-reduction method in which linear kernels are used for the nonlinear mapping of input space into feature space. On the other hand, the projection of data onto selected eigenvectors in kernel PCA is equivalent to performing CMDS on the transformed dissimilarities when the kernel functions are Gaussian. Thus nonlinear dimensionality reduction using CMDS can be obtained by using reproducing kernels as similarities in lieu of Euclidean distance (Williams, 2002; Zhang & Jordan, 2009). In this study, a nonlinear multivariate statistical process monitoring framework is used based on the equivalence between reproducing kernels and inner products (similarities), building on top of a linear approach using CMDS that has recently been used for industrial process monitoring (Yunus & Zhang, 2014).

2.2.5 Kernel multidimensional scaling (kernel MDS)

For matrix $\mathbf{X}$ with a set of n points and m variables, the $n \times n$ dissimilarity matrix,$\mathbf{D}$, is constructed based on the Euclidean distance between variables i and j, d_{ij}, which is calculated using equation (2.2.40).

The squared dissimilarity matrix is double centered to compute the inner product matrix $\mathbf{B}$:

$$\mathbf{B} = -\frac{1}{2}\,\mathbf{J}_n\mathbf{D}^2\mathbf{J}_n = \mathbf{J}_n\mathbf{X}\mathbf{X}^T\mathbf{J}_n \qquad \textbf{(2.2.47)}$$

where $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n]$, $\mathbf{J}_n = (\mathbf{I}_n - \mathbf{1}_n\,\mathbf{1}_n^T/n)$, $\mathbf{I}_n$ is an identity matrix, and $\mathbf{1}_n$ is a vector of length n with all elements equal to 1.

In order to establish the relationship between the dissimilarity matrix in (2.2.40) and the reproducing kernels, the positive definite kernel function $\boldsymbol{K}$ is used to compute the inner products in the feature space $\mathcal{F}$, which is related to the input space $\mathbf{F}$ by a nonlinear map,

$\boldsymbol{\phi}$. That is, the kernel function can be used to compute the value of inner products in $\mathcal{F}$ without carrying the nonlinear map $\boldsymbol{\phi}$ and is represented by

$$\boldsymbol{K}(\mathbf{f}_i, \mathbf{f}_j) = \boldsymbol{\phi}(\mathbf{f}_i)^T \boldsymbol{\phi}(\mathbf{f}_j) \tag{2.2.48}$$

where $\mathbf{f}_i$ and $\mathbf{f}_j$ are the input vectors in $\mathbf{F}$.

Gaussian kernels are used to obtain the squared (Euclidean) distance, d_{ij}^2 , between the vectors in the feature space, as shown in the following equation:

$$d_{ij}^2 = \boldsymbol{K}_{ii} + \boldsymbol{K}_{jj} - 2\boldsymbol{K}_{ij} \tag{2.2.49}$$

where $\boldsymbol{K}_{ij} = \boldsymbol{K}(\mathbf{f}_i, \mathbf{f}_j)$. Thus, in this study, the double-centred inner product matrix $\mathbf{B}$ can be expressed in terms of kernel function as

$$\mathbf{B} = \mathbf{J}_n \boldsymbol{K} \mathbf{J}_n \tag{2.2.50}$$

where each entry of the linear kernel matrix, $\boldsymbol{K}$, is the dot product between the Euclidean vector, $\boldsymbol{K}_{ij} = \mathbf{x}_i^T \mathbf{x}_j$. Any type of inter-distance can be transformed into $\mathbf{B}$, based on the relationship between the dissimilarity matrix $\mathbf{D}$ and the kernel matrix $\boldsymbol{K}$. The eigendecomposition of $\mathbf{B}$ in equation (2.2.50) approximates the feature space distance. Thus the dimensionality reduction of data in the feature is obtained by projecting them onto the first few eigenvectors; this is equivalent to performing CMDS in the feature space. This procedure is called *kernel MDS* (Williams, 2002).

Given $\mathbf{B}$, either by equation (2.2.47) or (2.2.50), CMDS finds the reconstructed coordinates of points by applying eigenvalue decomposition on $\mathbf{B}$, and hence $\mathbf{B}\mathbf{v}_i = \lambda_i \mathbf{v}_i$. That is,

$$\tilde{\mathbf{X}} = \mathbf{V}_p \sqrt{\boldsymbol{\Lambda}_p} \tag{2.2.51}$$

where $\boldsymbol{\Lambda}_p$ is the diagonal matrix with p-ordered eigenvalues, and $\mathbf{V}_p$ is the $n \times p$ matrix whose columns correspond to the first p eigenvectors of $\mathbf{B}$. Thus the reconstructed point is $\tilde{\mathbf{x}}_i = \sqrt{\lambda_i}\mathbf{v}_i$.

In KPCA, the sample covariance matrix of vectors, $\mathbf{f}_i$, in the feature space is used to extract the nonlinear principal scores, which are given by

$$\mathbf{C} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{f}_i - \bar{\mathbf{f}})(\mathbf{f}_i - \bar{\mathbf{f}})^T = \frac{1}{n}\mathbf{F}^T\mathbf{J}_n\mathbf{F} = \frac{1}{n}\mathbf{F}^T\mathbf{J}_n\mathbf{J}_n\mathbf{F} = \frac{1}{n}(\mathbf{J}_n\mathbf{F})^T(\mathbf{J}_n\mathbf{F}) \qquad \textbf{(2.2.52)}$$

where $\bar{\mathbf{f}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{f}_i$. Since $\mathbf{F}^T\mathbf{J}_n\,\mathbf{J}_n\mathbf{F}$ has the same nonzero eigenvalues as $\mathbf{J}_n\mathbf{F}\,\mathbf{F}^T\mathbf{J}_n = \mathbf{J}_n\boldsymbol{K}\,\mathbf{J}_n = \mathbf{B}$, it can be shown that the eigenvalues of $n\mathbf{C}$ are the p nonzero eigenvalues of $\mathbf{B}$ in KPCA. To show this, note that $\mathbf{J}_n^2 = \mathbf{J}_n$, and thus that $n\mathbf{C} = (\mathbf{J_n}\mathbf{F})^T(\mathbf{J}_n\mathbf{F})$. Let $\mathbf{v}_i$ be a unit-length eigenvector of $\mathbf{B}$, so that $\mathbf{B}\mathbf{v}_i = \lambda_i\mathbf{v_i}$. Premultiplying by $(\mathbf{J}_n\mathbf{F})^T$ gives

$$(\mathbf{J}_n\mathbf{F})^T(\mathbf{J}_n\mathbf{F})(\mathbf{J}_n\mathbf{F})^T\mathbf{v}_i = \lambda_i(\mathbf{J}_n\mathbf{F})^T\mathbf{v}_i \qquad \textbf{(2.2.53)}$$

where λ_i is an eigenvalue of $n\mathbf{C}$, and $\mathbf{e}_i = (\mathbf{J}_n\mathbf{F})^T\mathbf{v}_i$ is the corresponding eigenvector in which $\mathbf{e}_i\mathbf{e}_i^T = \lambda_i$. Centring $\mathbf{F}$ and projecting it onto the unit vector, $\tilde{\mathbf{e}}_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{e}_i$, obtains the p-dimensional configuration of $\mathbf{F}$. That is

$$(\mathbf{J}_n\mathbf{F})\tilde{\mathbf{e}}_i = \frac{1}{\sqrt{\lambda_i}}(\mathbf{J}_n\mathbf{F})(\mathbf{J}_n\mathbf{F})^T\mathbf{v}_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{J}_n\mathbf{F}\mathbf{F}^T\mathbf{J}_n\mathbf{v}_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{B}\,\mathbf{v}_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{v}_i\,. \qquad \textbf{(2.2.54)}$$

Thus we see that the projection of $\mathbf{F}$ onto the eigenvalues of $n\mathbf{C}$ returns the classical scaling solution. This proves the duality between KPCA and CMDS.

CMDS can be performed in the feature space by using different kernel functions. For example, the use of linear kernels gives the basic CMDS, because $\mathbf{f}_i = \mathbf{x}_i$. Hence CMDS can be used as an alternative dimensionality-reduction technique because there exist a nonlinear mapping from $\mathbf{f}_i$ to $\mathbf{x}_i$ (Zhang & Jordan, 2009).

In summary, kernels can be considered as generalised dissimilarity measures that are positive definite, and can be viewed as a nonlinear generalisation of the similarity measure based on inner products. Thus a new framework for the distance-based algorithm is devised by using suitable kernel functions for the realisation of squared Euclidean distance in the feature space (Pekalska *et al*., 2001; Schölkopf, 2000). This motivated the use of the kernel MDS algorithm in this study as a form of performing CMDS in feature space using Gaussian kernels.

Process monitoring using MDS was first proposed by Cox (2001), who summarised the multivariate data structure by means of conventional score configuration (observation-sample) profiles. Also, the occurrence of an abnormal event results in a deviation of the sample in greater magnitude from the normal sample. In another approach by Cox (2003), the status of the monitoring performance was introduced by a variable-based form of scores in which the variables responsible for fault conditions not only show the deviation, but also project the information on the deviation of the normal variable's coordination in terms of multivariate scores. A multiple linear regression was used in projecting the MDS scores for online monitoring and fault detection by Matheus *et al*. (2006) based on topographic mapping and clustering operation in another study. However, the above studies do not propose any kind of monitoring statistics as the basic measure for fault detection. Three new process monitoring frameworks using CMDS as an alternative tool for dimensionality reduction based on the dissimilarity measure in describing the correlation between the process variables were proposed recently by Yunus and Zhang (2010). Thus the limitations of conventional multivariate methods for monitoring nonlinear processes are addressed. A reduced dimensional space for an effective process monitoring scheme was obtained on the basis of variable scores instead of sample scores as in PCA (Yunus & Zhang, 2010). Even though fault detection was improved in these studies, there were difficulties in understanding the real effect of different inter-distance scaling on the fault detection performance between the process systems. A new MSPM framework has been proposed by Yunus and Zhang (2014), in which the correlation among the samples are measured using dissimilarity scale structure and $\boldsymbol{T^2}$ and SPE statistics for fault detection. Hence the original conceptual framework of MSPM is maintained to a great extent, without focusing too much on new terminologies, as was the cases in previous studies. This MSPM framework is used as an alternative monitoring tool to address the issues with PCA in this study.

2.2.6 Multiscale process monitoring

While traditional PCA and its various extensions assume that the chemical processes operate at a single-scale steady condition, the operating conditions of modern processes vary from time to time or move from one operating condition to another. The operating conditions of processes in practical situations change over time because of the variations in the availability of raw materials, changes in set-point, equipment degradation and the effects of environmental (seasonal) changes. Nevertheless, the

process personals have to meet the market demands for product quality and its safe operation. Hence the application of a stable monitoring approach to such operations may give rise to a false alarm, even when the process operates in another steady-state nominal mode. In order to keep the process under control, the model has to be updated automatically or adaptive models must be used (Sun *et al*., 2012). In the past years, another monitoring technique based on wavelet transforms has been developed to handle the changes in the process condition. Multiscale PCA (MSPCA), a combination of wavelet transform and PCA, was introduced into the process monitoring scheme by Bakshi (1998) to separate the components in the data into multiple time scales. When the data is decomposed into several time scales, the separated time scale is indirectly close to having a constant mean, which solves the issues when PCA is applied (Mirin and Wahab, 2014).

In multiscale PCA, wavelets are used to decompose the data into several views or scales before the application of PCA to detect the faults in the process operations. Here, the ability of PCA to decorrelate variables is combined with the ability of wavelets to decompose the data into multiple scales, eliminating autocorrelation in the observations due to the time-frequency localisation property of wavelets. In order to monitor the multiscale process, several combinations of PCA with wavelets have been developed because of the ability of wavelets to separate multiscale features and approximately remove serial or auto-correlations in time signals (Bakshi, 1998; Maulud *et al*., 2006; Misra *et al*., 2002; Rosen & Lennox, 2001). The MSPCA approach has been extended to monitor the nonlinear process by using neural networks to extract the latent nonlinear structure from the PCA-transformed data (Fourie & De Vaal, 2000; Shao *et al*., 1999; Geng & Zhu, 2005). Shao *et al*. (1999) proposed a nonlinear process monitoring and fault diagnosis method for the application of an industrial process based on wavelets and NLPCA, using an input-training neural network to extract both linear and nonlinear correlation from the process data in which the data is first pre-processed using wavelets to remove noise and spikes prior to the application of NLPCA in the monitoring algorithm. Fourier and De Vaal (2000) also developed a nonlinear multiscale principal component analysis (NLMSPCA) technique for process monitoring and fault detection based on multilevel wavelet decomposition and NLPCA, using input-training neural networks, in which the multiscale representation of the data obtained by using wavelets enhances the performance of the monitoring scheme to detect different types of

abnormal conditions in the process. Geng and Zhu (2005) presented a novel method of wavelet-based adaptive multiscale nonlinear PCA (MS-NLPCA), using an improved input-training neural network (IT-NN) for monitoring process signals. The slow and feeble changes in fault signals are effectively and accurately detected with a minimum rate of false alarms. A multiscale orthogonal nonlinear strategy for multivariate statistical process monitoring was proposed by Maulud *et al.* (2006), using optimal wavelet decomposition and an orthogonal NLPCA algorithm to capture the nonlinear characteristics of the process data with a minimum number of principal components.

The concept of KPCA has also been combined with multiresolution analysis for nonlinear multiscale process monitoring. In the nonlinear multiscale modelling strategy for fault detection and identification proposed by Choi *et al.* (2008), they applied KPCA in the model built with the reconstructed data obtained by performing wavelet transform and inverse wavelet transform sequentially on the process data. Deng and Tian (2006) developed a monitoring method based on multiscale KPCA for nonlinear dynamic processes in which wavelet analysis is combined with nonlinear transformation using KPCA. Zhang and Ma (2011) proposed a novel nonlinear process monitoring and fault diagnosis technique called multiscale KPCA, and multiscale kernel partial least squares based on KPCA and kernel partial least squares (KPLS) models at different scales. In the proposed monitoring algorithm, KPCA and KPLS are applied to the multiscale data obtained by wavelet decomposition to capture the correlation of process variables at different scales. In the abovementioned multiscale methods based on wavelets and KPCA, wavelets are used to analyse the dynamic characteristics of process data, while the KPCA is used to capture nonlinear principal components by kernel function.

Although wavelets can extract deterministic and stochastic features at various scales, the success thereof is dependent on a number of factors, the most significant being the choice of a basis function or mother wavelet for the optimal orthogonal expansion of the signal for the application at hand. This mother wavelet should also have some desirable properties, such as good time frequency localisations and general admissibility properties, including various degrees of smoothness (number of continuous derivatives) and a large number of vanishing moments (ensures maximum number of zeros of the polynomial at the highest discrete frequency) (Daubechies, 1992; Ganesan *et al*., 2004; Meyer, 1992). A large number of wavelet bases have been developed to meet the

requirements, such as completeness, time-frequency localisation, and orthogonality or limited redundancy for non-orthogonal base functions. Clearly, choosing an appropriate wavelet basis function for a specific purpose is a difficult task for practitioners. In addition, while the optimal multiscale decomposition of the signal can be obtained by an automatic time-varying adjustment in the mother wavelet's shape, it does not allow for adjusting the nature of the analysing function that is adaptable to the signal (Krishnannair, 2010).

For nonlinear and non-stationary data, an adaptive basis function that is derived from the data is more appropriate to represent the variety of underlying physical components in the process. Unfortunately, wavelet-based multiscale methods have a prior basis and they are not adaptive. Thus, to overcome the above limitations of wavelet-based multivariate methods, a data-adaptive multimodal method adapted to the nature of data, such as singular spectrum analysis (SSA), was developed and nonlinear versions of SSA are proposed for handling the nonlinearity and non-stationary behaviour of the process in the following chapters.

CHAPTER 3: PROCESS MONITORING USING SINGULAR SPECTRUM ANALYSIS

Singular spectrum analysis (SSA) has recently become a promising tool to extract information from short and noisy auto-correlated time series by decomposing the data into deterministic and stochastic components without prior knowledge of the dynamics affecting the time series. These components can be classified as independent additive time series of slowly varying trends, periodic series and aperiodic noise (Golyandina *et a*l., 2001). SSA does this decomposition by projecting the original time series onto a data-adaptive vector basis obtained from the series itself, based on principal component analysis (PCA). A multiscale SSA strategy has been proposed to capture both the time and frequency domain by treating each process variable as a time series, and the auto-correlation between the variables is explicitly accounted for (Krishnannair, 2010). It is found that, in most cases, the proposed method is superior in detecting the gradual or slow process changes and faults of different magnitudes accurately, compared to classical statistical process control (SPC) based on latent variable methods. SSA uses data-adaptive basis functions and therefore can be expected to provide more flexibility than other spectral techniques.

Singular spectral time-series analysis pre-filters the original time series into a sum of different components such as trend, periodic or quasi-periodic, and noise. This is done by the singular value decomposition (SVD) of a trajectory or lagged covariance matrix obtained from the original time series, followed by reconstruction of the series using subsets of eigenfunctions and corresponding principal components.

More specifically, the SSA methodology is summarised as the following procedure, the first step of which is the embedding of the time series into a M-dimensional time series known as the *trajectory matrix*. In the second step, singular value decomposition is applied to decompose the trajectory matrix into a sum of elementary matrices.

Subsequently, the elementary matrices that contribute to the norm of the original matrix are grouped, with each group giving an approximation of the original matrix. Finally, the smoothed approximation of the time series is recovered by diagonal averaging of the elementary matrices obtained from decomposing the trajectory matrix.

3.1 SSA METHODOLOGY

An outline of the basic SSA methodology is summarised in Figure 3.1 (Golyandina *et al*., 2001), and the procedural steps involved in SSA are adopted from Jemwa and Aldrich (2006) and Krishnannair (2010).

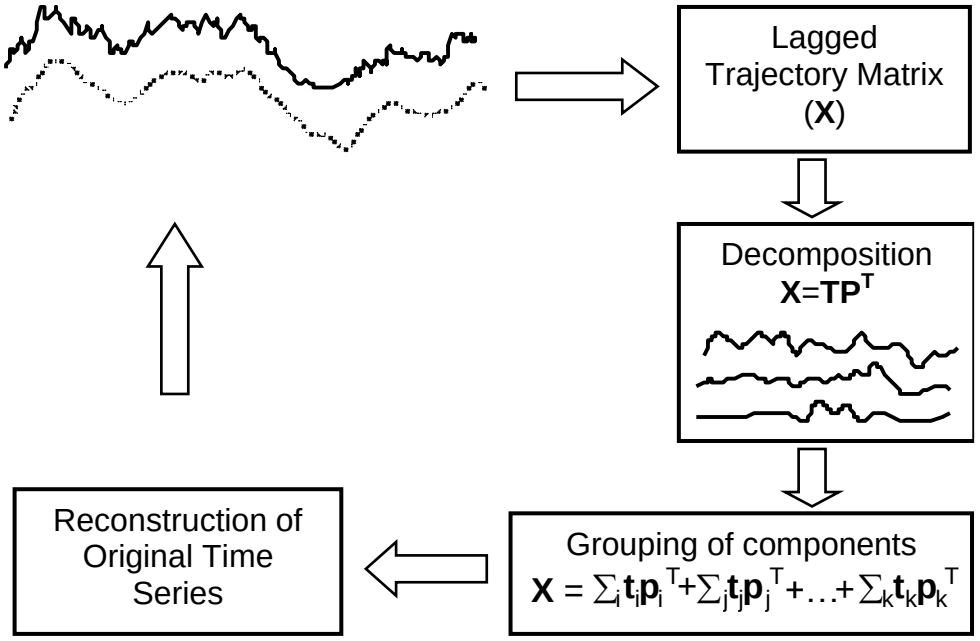


Figure 3.1: Embedding, decomposition, grouping and reconstruction of a time series usingf singular spectrum analysis.

3.1.1 Step 1: Embedding

In this step, a time series, $x(t)$, $t = 1,2,\dots,N$, of length N is embedded with a window of length M to construct K lagged vectors, $\mathbf{x}_i \in \Re^M$:

$$\mathbf{x}_i = (x(i), x(i+1), \dots, x(i+M-1))^T \ , \ 1 \le i \le K, \ K = N - M + 1. \qquad \textbf{(3.1.1)}$$

These embedded vectors are then augmented into a multidimensional time series, commonly referred to as the trajectory matrix in the study of nonlinear dynamic systems, since it represents a trajectory of the evolution of the dynamical system represented by

the observed time series $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K)^T$ (Broomhead & King, 1986; Ghil *et al.*, 2002):

$$\mathbf{X} = \left(x_{ij}\right)_{i,j=1}^{K,M} = \begin{bmatrix} x(1) & x(2) & \cdots & x(M) \\ x(2) & x(3) & \dots & x(M+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(K) & x(K+1) & \dots & x(N) \end{bmatrix}. \quad \textbf{(3.1.2)}$$

The augmented matrix, $\mathbf{X} \in \Re^{K \times M}$, is in the form of a a Hankel matrix (Strang, 2009), that is,

$$x_{ij} = x(i+j-1), \;\; 1 \le i \le K, \;\; 1 \le j \le M.$$

3.1.2 Step 2: Singular value decomposition

A $M \times M$ covariance matrix, $\mathbf{C_x}$ of , $\mathbf{X}$ is constructed from the trajectory matrix, i.e.

$$\mathbf{C_x} = \frac{1}{K}\mathbf{X}^T\mathbf{X} \quad \textbf{(3.1.3)}$$

where $K = N - M + 1$, and the eigendecomposition of $\mathbf{C_x}$ is computed as a solution of the eigenvalue problems

$$\mathbf{C_x}\mathbf{a}_k = \lambda_k \mathbf{a}_k \,, k = 1,2, \dots, M \quad \textbf{(3.1.4)}$$

where $\mathbf{a}_k$ and λ_k are the k^{th} eigenvector and eigenvalue respectively.

The square root of the non-negative eigenvalues $\left(\sqrt{\lambda_k}\right)$ are called the singular values, and the ordered set of singular values, $\sqrt{\lambda_1} \ge \sqrt{\lambda_2} \ge \dots \ge \sqrt{\lambda_M}$, is called the *singular spectrum*, from which SSA derives its name. The ordering implies that the k^{th} eigenvalue explains at least as much of the variance in the data as the $(k+1)^{th}$ eigenvalue. Broomhead and King (1986) expressed the singular value decomposition (SVD) of the trajectory matrix in terms of eigenfactors of $\mathbf{C_x}$, by setting $d = \max\{k \in \{1,2,\dots,M\}/\lambda_k > 0\}$ and $\mathbf{v}_k = \mathbf{X}^T\mathbf{a}_k/\sqrt{\lambda_k}$ for $k = 1,2,\dots,d$, and it can be written as

$$\mathbf{X} = \sum_{k=1}^{d} \sqrt{\lambda_k}\, \mathbf{a}_k\mathbf{v}_k^T = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_d \quad \textbf{(3.1.5)}$$

where $\mathbf{X}_k = \sqrt{\lambda_k}\,\mathbf{a}_k\mathbf{v}_k^T$ are bi-orthogonal matrices with rank one, known as elementary matrices.

An alternative approach estimates the covariance matrix $\mathbf{C_x}$ directly from the data such that its entries $\mathbf{C_x}(i,j)$ depend only on the lag $|i-j|$, that is, as a Toeplitz matrix (Elsner & Tsonis, 1996). Advantages of using each of the approaches have been discussed in the literature (Elsner & Tsonis, 1996; Ghil *et al*., 2002). Without loss of generality, the SVD approach is used in this study.

The SVD of the trajectory matrix in equation (3.1.5) can also be written as a product of a score matrix, $\mathbf{T} \in \Re^{K\times M}$, and a transposed loading matrix, $\mathbf{P} \in \Re^{M\times M}$, similar to PCA. That is,

$$\mathbf{X} = \mathbf{T}_i\mathbf{P}_i^T = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \cdots + \mathbf{t}_d\mathbf{p}_d^T \qquad \textbf{(3.1.6)}$$

where $\mathbf{P}_i = \sqrt{\lambda_i}\mathbf{v}_i$ and $\mathbf{T}_i = \mathbf{a}_i$.

Thus, the mathematical and statistical properties of PCA extend to SSA, since SSA is PCA performed on the trajectory matrix. When the variables are highly correlated, the leading first few principal components (PCs) capture most of the information in the data. Moreover, the approximation error can be minimised by representing the data by the first few leading PCs. When the data is distributed normally, the first few leading PCs have minimal entropy. For monitoring purpose, a PC explaining a larger portions of the variance in the data is considered (Aldrich & Barkhuizen, 2003; Jemwa & Aldrich, 2006; Krishnannair, 2010).

3.1.3 Step 3: Grouping of components

The additive components of the time series can be separated into the sum of intrinsic dynamical components and external noisy components through grouping. A large part of the information for the signal can also be compressed through grouping by projecting the time series on the subspace of the PCs corresponding to the largest singular values (Vitanov *et al*., 2008).

Each eigenvector obtained from the eigenvalue decomposition can be used to construct a time series of length $K = N - M + 1$ by projecting the vectors along each principal direction, k*:*

$$\mathbf{t}_k(t) = \sum_{j=1}^{M} x(t+j-1)\mathbf{a}_k(j) \qquad \textbf{(3.1.7)}$$

for $t = 1,2,\dots,K$ to obtain principal component scores $\mathbf{t}_k$. These scores represent the new coordinates of the data in the rotated coordinate space. Hence, the time series is represented by $p < M$ leading components, and the p-dimensional score vectors of the decomposed matrix, $\mathbf{T}$, are given by

$$\mathbf{t}(t) = \left[\mathbf{t}_1(t), \mathbf{t}_2(t), \dots, \mathbf{t}_p(t)\right]^T,\ t = 1,2,\dots,K\ . \qquad \textbf{(3.1.8)}$$

The selection of leading components to retain for the extraction of signal features such as trends and oscillations can be done in many ways. For example, grouping the indices from slowly varying eigenvectors can extract a trend and oscillations can be extracted by grouping those indices from eigenvector pairs whose scatter plot resembles a circle or a polygon (Thomakos *et al.*, 2002). The percentage contribution of the k^{th} principal component in the analysed time series can also be used to group components to retain. The percent contribution of an eigenvector (or fraction of explained variance) is represented by the ratio (Tzagkarakis *et al*., 2007)

$$R_k = \frac{\lambda_k}{\sum_{i=1}^{M}\lambda_j}\ . \qquad \textbf{(3.1.9)}$$

The contribution of the elementary matrix, $\mathbf{X}$, to the expansion of the trajectory matrix, $\mathbf{X}$, can also be represented by the above ratio (Alonso & Salgado, 2005).

3.1.4 Step 4: Reconstruction

The convolution of a set of principal components,$\mathbf{T}$, with corresponding eigenvector or principal directions, recovers phase information lost in the preceding decomposition:

$$\tilde{\mathbf{x}}(t+j-1) = \sum_{k=1}^{p\leq M}\mathbf{t}_k(t)\mathbf{a}_j(k),\ t = 1,2,\dots K, j = 1,2\dots,M\ . \qquad \textbf{(3.1.10)}$$

The approximation of the reconstructed time series can be obtained by taking the average of the elements on the corresponding diagonals of the elementary matrices obtained in the grouping stage. Through diagonal averaging, or *Hankelization*, the elementary matrix obtained in equation (3.1.10) is transformed into a principal component of length *N* to create reconstructed components (RCs) of the original series. The diagonal averaging is performed according to

$$\tilde{\mathbf{x}}(i) = \begin{cases} \frac{1}{i}\sum_{j=1}^{i}\sum_{k=1}^{p}\mathbf{t}_k(i-j+1)\mathbf{a}_k(j) & 1 \le i \le M-1 \\ \frac{1}{M}\sum_{j=1}^{M}\sum_{k=1}^{p}\mathbf{t}_k(i-j+1)\mathbf{a}_k(j) & M \le i \le N-M+1 \\ \frac{1}{N-i+1}\sum_{j=i-N-M}^{M}\sum_{k=1}^{p}\mathbf{t}_k(i-j+1)\mathbf{a}_k(j) & K+1 \le i \le N \end{cases} \quad \textbf{(3.1.11)}$$

where $K = N - M + 1$.

The diagonal averaging on the elementary matrices leads to the decomposition of the original series into p reconstructed components of the fitted values of the reconstruction and residual series, capturing the error in the reconstruction. In most physical applications, the fitted values of the reconstruction are associated with deterministic components (trends) and the residual series is associated with stochastic components (noise) in the data. Hence the reconstruction step separates the underlying signal from the noise (Thomakos *et al*., 2002). Moreover, the reconstruction step preserves the phase of the time series so that $x(t)$ and $\tilde{x}(t)$ can be superimposed on the same time scale, $1 \le t \le N$. This is an advantage of using RCs over PCs of length K, as these do not contain direct phase information within the embedding dimension M (Krishnannair, 2010). The reconstructed components, $\tilde{x}(i)$, recover phase information of the time series lost in the decomposition stage, and the diagonal averaging in equation (3.1.11) is an adaptive optimal filter in the least-squares sense (Ghil *et al.*, 2002; Vautard *et al*., 1992).

The typical decomposition of a signal using SSA is illustrated in Figure 3.2. The original univariate, auto-correlated and noisy time series is plotted in Figure 3.2(a). The decomposition of the original data into five different modes is achieved by using the SSA algorithm, as described above. An embedding window of length five was used to build the trajectory matrix. The high frequency features in modes 2 to 5 are plotted in Figure 3.2(c-f). The low-frequency features in the first mode are plotted in Figure 3.2b. The low-frequency signals in the first mode are reconstructed by using those PCs with the largest eigenvalues, and the remaining components are reconstructed by using those PCs with eigenvalues decreasing in the ascending order. SSA decomposes the signal into principal signals that capture distinct features in the original signal, namely the deterministic mean and random variations (Tzagkarakiz *et al*., 2007). The first few deterministic components carry most of the information content of the original signal and can give more accurate predictions because of their slow-varying nature. The reconstruction of signals at different modes in Figure 3.2 is obtained by using a data-

adaptive basis function, which has been adapted to the different features of the signal (Krishnannair, 2010).

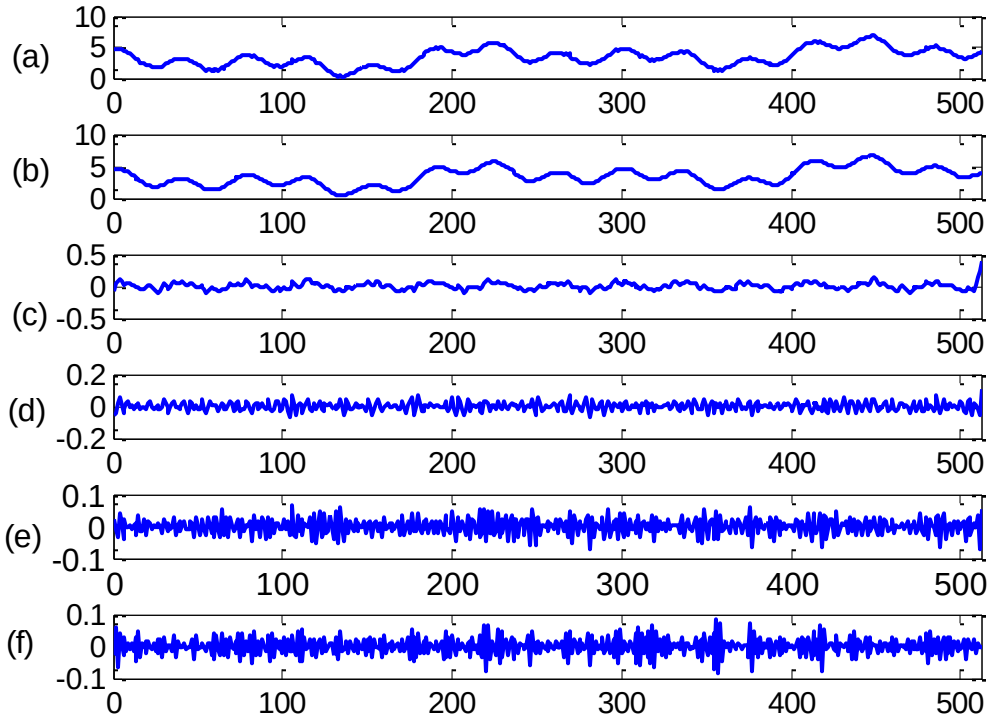


Figure 3.2: Decomposition of an (a) observed signal into multiple scale representations using singular spectrum analysis. Here, a sliding window M = 5 was used. The significant or deterministic component (b), explaining the most variation in the signal, is associated with the leading eigenelement, with (c) to (f) progressively explaining less variability. In particular, the last signal is associated with high frequency components (Krishnannair, 2010).

Applications of SSA have been widely reported in climatology and geophysical sciences, which include the analysis of paleoclimatic time series for the detection of climatic oscillations and regime changes in amplitude (Vautard & Ghill, 1989); the analysis of global surface air temperature time series for the extraction of global warming trends and oscillatory modes from the noisy components (Vautard *et al*., 1992); climatic oscillations reported in the Guliya ice core, their connections to variations in solar radiation and the recorded sub-orbital climatic oscillations (Yang *et al*., 2006); and the detection of periodicities and trends in the temporal dynamics of soil salinity (Florinsky *et al*., 2009).

SSA has also been used for the extraction of relevant trends and oscillations in the time series in a comparative study on climatic oscillations (Aldrich & Barkhuizen, 2003; Ghil & Yiou, 1996).

Applications of SSA in the biosciences have also been reported in certain studies. Mineva and Popivanov (1996) investigated the dynamics of the single-trial readiness potentials in human beings using EEG (electroencephalogram) signals. While Pereira and Macial (2001) used SSA in ultrasonic analysis, Pereira *et al*. (2004) extended the study to the characterisation of the properties of trabecular bones. Other related applications of SSA include ultrasonic detection and imaging of brachytherapy seeds (Mamou & Feleppa, 2007); the detection of spatial and temporal variations in shoreline positions to identify characteristic patterns in the shoreline response (Rozynski *et al*., 2001); the analysis of time series with missing data points (Schoellhamer, 2001); and signal-to-noise ratio enhancement (Carniel *et al*., 2006). The use of SSA has also been reported in econometrics (Hassani & Thomakos, 2010; Hassani and Zhigljavsky, 2009; Hassani *et al*., 2009; 2010; 2013; Thomakos *et al*., 2002;); biomechanical analysis (Alonso *et al*., 2005); machine condition monitoring (Alonso & Salgado, 2005; Salgado & Alonso, 2006; 2007; Wang *et al*., 2001); computer network behavioural analysis (Tzagkarakis *et al*., 2007; Wu & Gong, 2000); safety control and monitoring in nuclear power plants (Palomo *et al*., 2003); pre-filtering/denoising and smoothing of data (Oropeza & Sacchi, 2011); system identification in metallurgical reactors (Aldrich & Barkhuizen, 2003); trend extraction (Alexandrov, 2009); time series classification (Jemwa & Aldrich, 2006); multimodal process monitoring (Krishnannair, 2010); and process analysis and performance assessment of sheet-forming processes (Yuan, 2015).

Since SSA is the principal component analysis performed on a trajectory matrix, all mathematical and statistical properties associated with PCA apply to SSA, as indicated before. The use of statistical concepts in the SSA framework does not require certain statistical assumptions, such as stationarity or the normality of residuals (Golyandina *et al.,* 2001; Hassani & Zhigljavsky, 2009; Vautard *et al*., 1989). SSA belongs to a family of methods based on empirical orthogonal function (EOF) expansion, whose main characteristic is the data-adaptive nature of the basis functions. The variance of the decomposition using EOF methods is defined in terms of the basis functions or mode

(i.e.$\{\mathbf{a}_k\}$ in equation (3.1.4)) and, therefore, unlike in wavelet decomposition, the variance distribution does not imply scales or frequency content of the signal. This limitation has been cited as a critical flaw of EOF-based methods (Huang *et al.*, 1998). However, because of the data-adaptive property, SSA and related methods have been used successfully in many applications, as discussed in the above paragraphs.

3.2 STATISTICAL PROCESS MONITORING USING SINGULAR SPECTRUM ANALYSIS

As discussed in the previous section, SSA separates a time series into additive components such as trend, anharmonic oscillations, or noise. Because of the data-adaptive nature of the basis functions, extracted anharmonic oscillations are usually expressed in terms of a much fewer number of the basis functions than would be required when using fixed basis functions, for example sines and cosines in Fourier analysis. Therefore, the separate analysis of extracted components may give information on the underlying dynamics of the physical system that cannot be extracted from the analysis or the use of raw measurements, possibly due to confounding influences such as noise, auto-correlation, or embedded features with different time-frequency localisations. This insight contributes to the development of multiscale process monitoring methods based on wavelet decomposition (Aradhye *et al*., 2003; Bakshi, 1998; Reis *et al*., 2008; Yoon & MacGregor, 2004). In these studies, the 'scale' represents the width of the scaling function or 'mother wavelet', and the wavelet coefficients at the same scale for different measurements can be considered collectively. The proposed use of SSA for process monitoring is motivated by similar reasoning, but exploiting the data-adaptive nature of the obtained basis functions in this section and in this study in general, and is named multiscale SSA (MS-SSA), as indicated in Chapter 1.

The basic framework for on-line and off-line process monitoring using SSA is illustrated in Figure 3.3. The input data can either be univariate or multivariate, with any of the following attributes: stationary or non-stationary; Gaussian or non-Gaussian; random or gross errors; independent or correlated; linear or nonlinear; deterministic or stochastic. In off-line training, the initial step is the decomposition of variables using SSA into different modes. Then the signal at each mode is subsequently reconstructed, preserving the phase of the original time series. After the reconstruction step, statistical process control (SPC) techniques such as PCA can be applied to monitor the

reconstructed data at different modes for significant information on faults, by computing the $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics in the respective modes. Finally, the control limits of the monitoring statistics are computed for fault detection. In on-line training, new sets of samples are used to determine the monitoring statistics. If the monitoring statistics exceed the control limits then the process is monitored to be out of control; otherwise it is normal. More detailed discussion of the above basic steps will follow in the discussion of the process monitoring method with SSA.

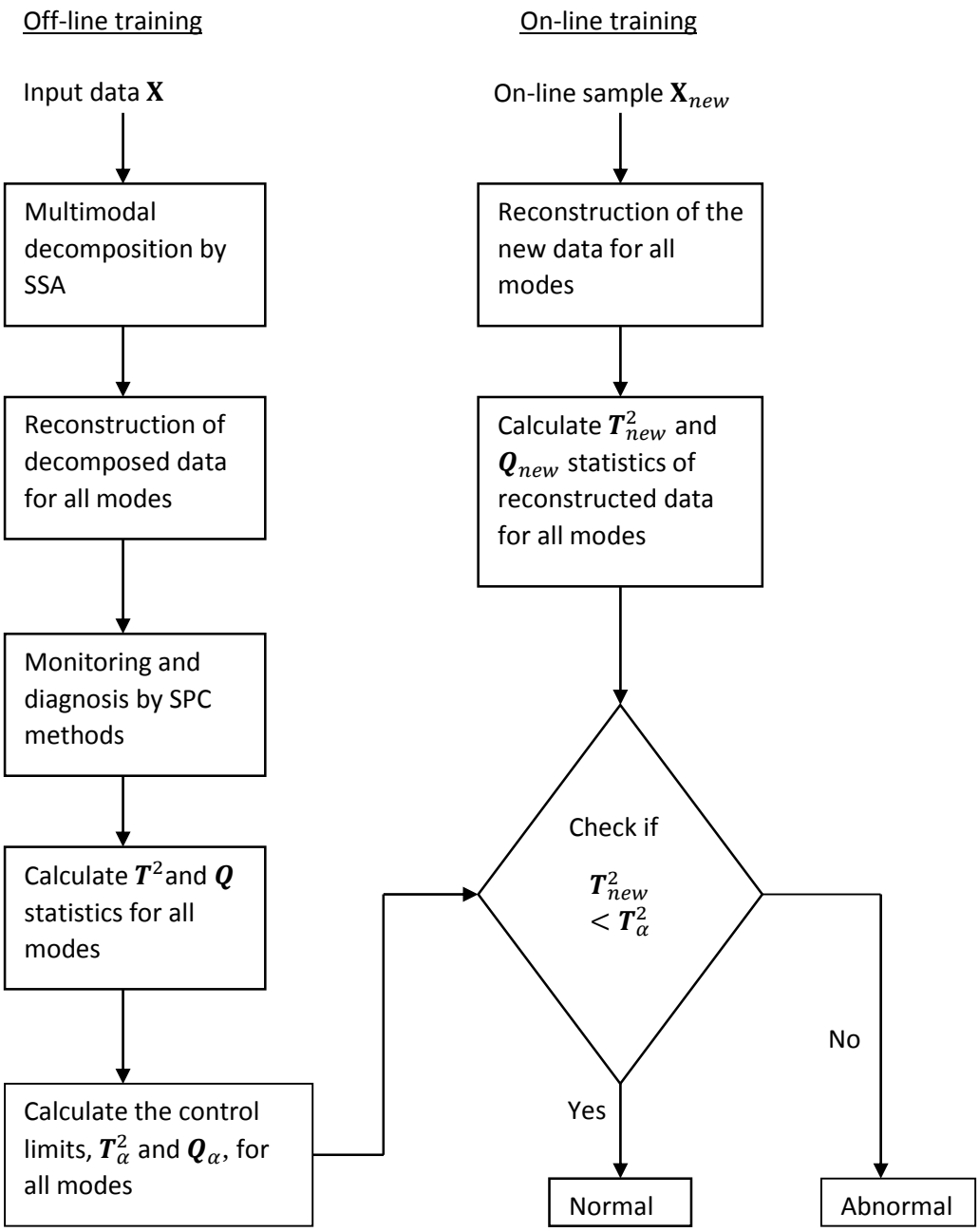


Figure 3.3: Basic steps in the statistical process monitoring method using SSA.

3.2.1 MS-SSA methodology

The process monitoring methodology using SSA, multiscale SSA (MS-SSA), involves the decomposition of each variable by SSA into multiple models, after which a PCA model is developed using the reconstructed variables of each mode, thereby accounting for the correlation between variables. Classical PCA is applied to each reconstructed multivariate data at each mode, representing the normal operation for calculating control

limits for scorers and residuals. For new data, a change in the normal operation (fault) is detected if the scores or residuals of the reconstructed data violate the control limits at any mode. Multimodal decomposition of the time series data using SSA is illustrated in Figure 3.4 and is discussed in detail below (Krishnannair, 2010).

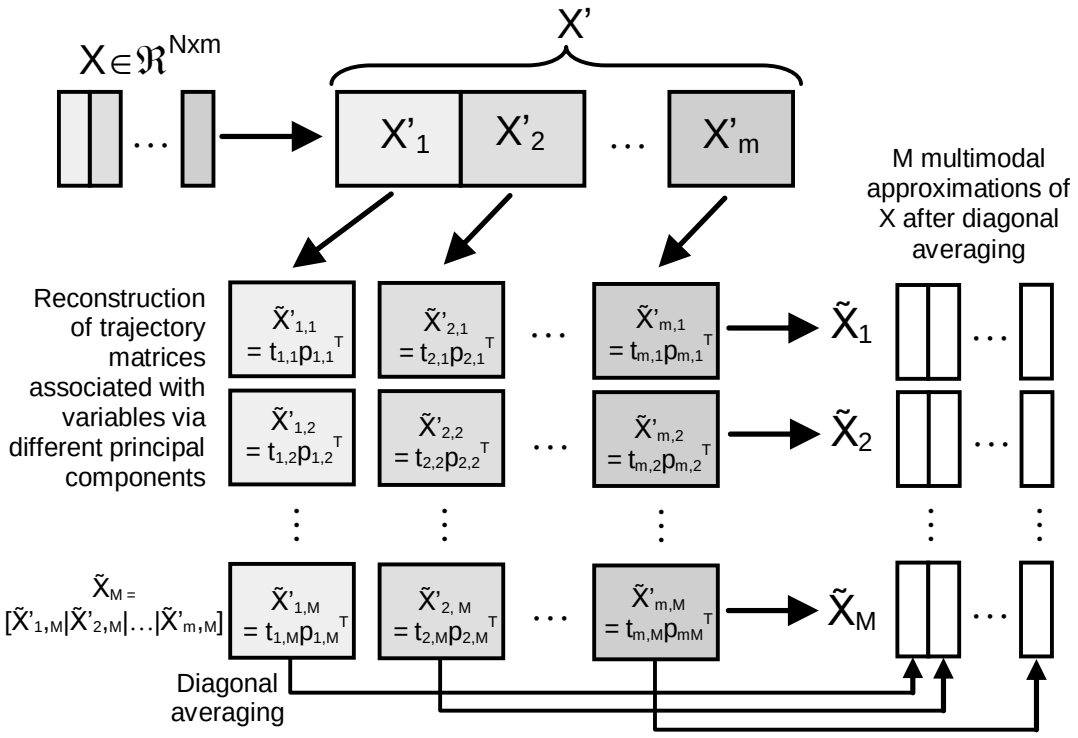


Figure 3.4: A schematic summary of multivariate SSA-based multimodal decomposition (Krishnannair, 2010).

3.2.1.1 Multimodal decomposition of data with SSA

In MS-SSA, the analysis of the process variables at different modes can be obtained by applying SSA to each variable separately, using a common window of size M. For a given multivariate time series data $\mathbf{X}$ with N observations on m variables, a trajectory matrix for each variable $\mathbf{X}'_k$ is computed by augmenting each variable with M lagged copies of itself using equation (3.1.2). Selecting the optimal M to use is a design challenge and its choice requires elaboration, as it is the parameter that controls the trade-off between the amount of significant information and the statistical confidence in the extracted information (Broomhead & King, 1986; Ghil *et al*., 2002).

Golyandina *et al.* (2011) discussed in detail the interplay between the choice of window size and the separability of features of interest. In general, large window sizes are preferable for a detailed decomposition of a time series, whereas a small window size allows for as many repetitions as possible of identical features. Poor choice of the window length may result in the mixing of interpretable components. Unfortunately, many time series are invariably disparate and, therefore, no general recommendation exist on the proper choice of window length. In practice, the choice of M is based mainly on heuristics, namely the first minimum of mutual information criteria and the first zero of the auto-correlation function (Abarbanel, 1997). The estimation of the window size using mutual information is based on the determination of a probability distribution on the system states and may result in a poor choice of the embedding dimension for any incorrect estimation of the probability distribution.

Estimation of the window size based on the first zero crossing of the auto-correlation function for the data depends on the linear independence between the two "state" variables, $x(t)$ and $x(t-M)$ (Bray & Wikswo, 2002). This method gives more reliable results than the mutual information criterion, because auto-correlation functions are useful for determining residuals, for detecting periodic components in data and for identifying the dominant power law noise type (white, flicker, random walk, flicker walk and random run) for the particular data type (Krishnannair, 2010). The selection of window length for SSA decomposition in this study was based on the first zero of the auto-correlation. Irrespective of what criterion was selected, the choice of window size must ensure that the lag is large enough for the coordinates to carry as much new information as possible, and small enough for the various coordinates not to be far apart. In this study, the choice of embedding window size based on the first maximum decorrelation point at which the sample auto-correlation was zero gave much more reliable results than what could be obtained by choosing some other point as decorrelation point for the auto-correlation function of the variables in the data.

The embedding dimension, M, for the data in the delay coordinate space can be selected such that the points of different lagged vectors in each variable, $\mathbf{x}_i(l), \mathbf{x}_i(k)$,($l \neq k$),$1 \leq i \leq m$, are linearly independent. In this study, the window length M was selected as the first maximal decorrelation point of variables, or the highest value of M where the sample auto-correlation function of each variable, $c_i(M)$, first crossed zero on

the y-axis (Bray & Wikswo, 2002). Mathematically, M is the $first.max\{M/c_i(M) = 0\}$, where

$$c_i(M) = \frac{\sum_{j=1}^{N}(x_i(j+M) - \bar{x}_i)\,(x_i(j) - \bar{x}_i)}{\sum_{j=1}^{N}(x_i(j) - \bar{x}_i)^2} \qquad \textbf{(3.2.1)}$$

and $\bar{x}_i = \frac{1}{N}\sum_{j=1}^{N} x_i(j)$ is the arithmetic mean of the corresponding variables $\mathbf{x}_i$.

Thus, given N observations of m variables, $\mathbf{X} \in \Re^{N\times m}$, each variable $\mathbf{x}_j$ for $j = 1,2,\ldots,m$ is decomposed by expressing its corresponding trajectory matrix in terms of an ordered series of score and loading vector products, i.e.

$$\mathbf{X}'_j = \mathbf{t}_{j,1}\mathbf{p}_{j.1}^T + \mathbf{t}_{j,2}\mathbf{p}_{j,2}^T + \cdots + \mathbf{t}_{j,M}\mathbf{p}_{j,M}^T \,. \qquad \textbf{(3.2.2)}$$

In SSA, the PCs with which the trend is obtained share some frequency bands with the PCs that represent the noisy component. SSA decomposes the signal without making any assumption about the frequency content of each PC, in contrast to other spectral techniques, which decompose the signal into disjoint frequency spectra (Golyandina *et al*., 2001; Krishnannair, 2010; Salgado & Alonso, 2006). In SSA, the spectrum of singular values only gives the proportion of variance explained by the PC directions and has no relation to the notion of scales or frequency of the signal. Hence, the individual principal directions will be referred to as modes and the statistical process control method can be considered *multimodal* as opposed to *multiscale*, as appearing in the study by Krishnannair (2010). The ranking of these modes is used to re-constitute the original multidimensional time series structure at multiple views or levels. There is no specific rule for selecting important modes for process monitoring, and the selection of significant modes is based mainly on heuristic criteria. However, important features can be extracted by monitoring the first few modes. Since PCs in the first few modes explain the maximum amount of variance in the data, those PCs are used to reconstruct the signal in the respective modes.

3.2.1.2 Reconstruction of data in all modes

The different components that represent the original signal in each mode are reconstructed by taking the average along the diagonals of each elementary matrix computed using equation (3.1.10).

Scaled versions of the original data matrix $\mathbf{X}$ are then reconstructed by

$$\tilde{\mathbf{X}}_i = \left[r(\mathbf{t}_{1,i}\mathbf{p}_{1,i}^T) \quad r(\mathbf{t}_{2,i}\mathbf{p}_{2,i}^T) \quad \dots \quad r(\mathbf{t}_{m,i}\mathbf{p}_{m,i}^T)\right] \tag{3.2.3}$$

where $r(.)$ is the diagonal averaging function (equation (3.2.3). The term $r(\mathbf{t}_{k,i}\mathbf{p}_{k,i}^T)$ represents the k^{th} reconstructed component obtained from the variable k for the i^{th} mode. Hence, at the coarsest level (mode $i = 1$), the data matrix $\mathbf{X}$ will be represented by all m process variables reconstructed from their respective first principal component loading and score vector pairs only, as shown in equation (3.2.4):

$$\tilde{\mathbf{X}}_1 = \left[r(\mathbf{t}_{1,1}\mathbf{p}_{1,1}^T) \quad r(\mathbf{t}_{2,1}\mathbf{p}_{2,1}^T) \quad \dots \quad r(\mathbf{t}_{m,1}\mathbf{p}_{m,1}^T)\right] . \tag{3.2.4}$$

This is repeated for all M modes, resulting in M representations of the data, as shown in Figure 3.4 above. More clearly, all the reconstructed components associated with the i^{th} principal directions or group in the decomposition are collected to form a multivariate series, $\tilde{\mathbf{X}}_i$, $i = 1,2,\dots,M$. In this sense, the method corresponds to multiscale methods based on wavelets, except that hierarchical representation is in the time domain and not in the wavelet domain.

3.2.1.3 Statistical process monitoring of the reconstructed signal

Once the multimodal representations have been obtained after decomposing and reconstructing each variable of the multivariate time series, SPC methods can be applied to each of the multiple-level representations for fault detection. Thus the M approximations of the original data are subsequently monitored separately using SPC methods. That is, each reconstructed matrix $\tilde{\mathbf{X}}_j$ is decomposed using PCA:

$$cov(\tilde{\mathbf{X}}_j)\tilde{\mathbf{p}}_{i,j} = \tilde{\mathbf{p}}_{i,j}\lambda_{i,j} \tag{3.2.5}$$

for $i = 1,2,\dots,M$ and $j = 1,2,\dots,m$, where $\lambda_{i,j}$ is the eigenvalue associated with the j^{th} eigenvector, $\tilde{\mathbf{p}}_{i,j}$, of the i^{th} representation of the original dataset.

Note that these loadings and corresponding scores, $\tilde{\mathbf{t}}_{i,j}$, are not the same as those used to approximate the original data at the different scales (since the matrices from which they are derived are different).

Similar to conventional PCA, the appropriate number of principal components retained (A) at each scale is selected and the control limits on the monitored indexes (Hotelling's $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics) are determined using a dataset obtained under normal operating conditions (e.g. Kresta *et al.*, 1991). Hotelling's $\boldsymbol{T}^2$ statistic for sample k is given by

$$\boldsymbol{T}_k^2 = \tilde{\mathbf{t}}_k \boldsymbol{\lambda}^{-1} \tilde{\mathbf{t}}_k = \tilde{\mathbf{x}}_k \tilde{\mathbf{P}}_A \boldsymbol{\lambda}^{-1} \mathbf{P}_A^T \tilde{\mathbf{x}}_k \qquad \textbf{(3.2.6)}$$

where $\tilde{\mathbf{t}}_k$ refers to the k^{th} row of $\tilde{\mathbf{T}}_A$, the matrix of A score vectors from the PCA model, and $\boldsymbol{\lambda}^{-1}$ is the diagonal matrix containing the inverse of the eigenvalues associated with the A principal components retained in the model. The squared residuals, or $\boldsymbol{Q}$ statistics, resulting from approximating the data by the PCA model can be shown to be

$$\tilde{\boldsymbol{Q}}_k = \tilde{\mathbf{e}}_k \tilde{\mathbf{e}}_k = \tilde{\mathbf{x}}_k (\mathbf{I} - \mathbf{P}_A \mathbf{P}_A^T) \tilde{\mathbf{x}}_k^T \qquad \textbf{(3.2.7)}$$

where $\tilde{\mathbf{e}}_k$ is the k^{th} row of $\tilde{\mathbf{E}}$, and $\mathbf{I}$ is the identity matrix of appropriate size.

Given new data, $\mathbf{X}_{new}$, a time-lagged expansion is performed on each variable using a window length of M. This is followed by eigenvalue decomposition of the resulting trajectory matrices. Subsequently, a set of multilevel representations of the original data is obtained using the same parameters as for the normal data. The values of the $\mathbf{T}^2$ and $\boldsymbol{Q}$ statistics are also calculated, using equation (2.2.12) and equation (2.2.13), for the new datasets and are compared to the corresponding control limits determined on the basis of the training set for each k^{th} representation. That is, the scores of the new data are calculated by projecting these data onto the A principal component loadings calculated with equation (3.2.3):

$$\tilde{\mathbf{t}}_{new,i,j} = \tilde{\mathbf{X}}'_{new,i} \tilde{\mathbf{p}}_{i,j}. \qquad \textbf{(3.2.8)}$$

If, at a specific mode, $\boldsymbol{T}^2$ or $\boldsymbol{Q}$ for the reconstructed new dataset is outside the calculated control limits, the process is judged to be out of control.

The use of multiple tests increases false positives error, i.e. the likelihood of incorrectly assigning an event as abnormal. For a set of independent tests, the significance level of each test must be adjusted such that the overall significance for all tests taken together equals the nominal value. Bonferroni's method can be used to adjust the significance

values (α) at each level (Bakshi, 1998; Yoon & McGregor, 2004). Thus, for an M level decomposition, the required Bonferroni adjustment is given by

$$\alpha_{adj} = 1 - (1 - \alpha_{nominal})^{1/M}. \quad \textbf{(3.2.9)}$$

In short, SSA is a nonparametric data adaptive technique and the application of SSA does not require any statistical assumption such as stationarity of an analysed series and normality of a residual. SSA can effectively handle stationary, non-stationary and seasonal time series. SSA decomposes a time series into interpretable components, such as trends, harmonic components and residuals for further applications such as forecasting and process monitoring. Since SSA can be applied to short time series and is beneficial for analysis of nonlinear dynamics (Hassani, 2007; Hassani & Zhigljavsky, 2009).

However, the multimodal SSA strategy that has been proposed to capture the features in multiple levels by treating each process variable as time series and the auto-correlation between the variables are explicitly accounted for (Krishnannair, 2010). It is found that in the case of multivariate auto-correlated systems, the proposed method is superior in detecting the gradual or slow process changes and faults of different magnitude accurately compared to classical statistical process control (SPC) based on latent variable methods. SSA uses data-adaptive basis functions and, therefore, can be expected to provide more flexibility than other spectral techniques.

CHAPTER 4: PROCESS MONITORING WITH NONLINEAR SINGULAR SPECTRUM ANALYSIS

In SSA, a linear approximation of the embedded dataset is sought, which accounts for the maximum amount of variance in the data. Since in SSA the lagged copies of time series are analysed by linear PCA, it can be expected that "*lossy*" feature extractors are obtained where a nonlinear description is more appropriate for the data on hand. Moreover, in SSA, the data is reconstructed based on the leading principal components and all other minor principal components are discarded in the reconstruction process. However, for data with nonlinear characteristics, the minor components always carry important information (Xu *et al.*, 1992). Unfortunately, retaining minor components has the undesirable effect of too many decomposition levels – a familiar variant of the *bias-variance trade off* scenario in statistical modelling (Hastie et al., 2011). SSA can also give misleading information in analysing anharmonic signals with too many modes since the signal energy is scattered into many SSA modes (Hsieh & Hamilton, 2003; Hsieh, 2004). Hence nonlinear extensions of SSA proposed by Hsieh and Wu (2002) have been used to deal with significant intrinsic nonlinearity associated with the real-world processes.

In this study, nonlinear generalisation of SSA, which is called multimodal nonlinear SSA, has been proposed with auto-associative neural networks (Kramer, 1191), classical multidimensional scaling (Cox & Cox, 1994) and kernel MDS (Williams, 2002) to exploit nonlinear redundancies that are detected by multimodal SSA. The proposed methods use nonlinear SSA to extract nonlinear features from the data by decomposing each variable into different modes as well as the nonlinear correlation of the variables. Thus, most of the limitations in the use of multimodal SSA in nonlinear data can be improved by evaluating the data at the respective modes with nonlinear singular spectrum analysis rather than basic SSA, as was the case in the previous studies.

Moreover, little work exists in the literature regarding the integration of MSPC with methods such as neural networks, CMDS, KMDS and singular spectrum analysis for process monitoring in chemical processes. Thus, in this study the aim was to improve monitoring performance of chemical process systems by the application of multivariate statistical methods that combine nonlinear singular spectrum analysis with the PCA monitoring method. Hence, an alternative, nonlinear, multimodal MSPC method is proposed to analyse the lagged copies of the time series in SSA that uses NLPCA, CMDS and kernel MDS instead of PCA in SSA decomposition. These methods are called multimodal NLSSA, multimodal DSSA and multimodal KDSSA. These methods aim to address some of the limitations of basic SSA-based process monitoring (MS-SSA) by introducing a variation in the SSA decomposition procedure. This is explained in detail in this chapter. In the multimodal NLSSA method (MM-NLSSA), NLSSA is used to extract the multilevel characteristics/features as well as the nonlinear correlation of variables in the data. The proposed MM-NLSSA combines the NLPCA concept with featured decomposition of data into multiple modes based on SSA. In multimodal DSSA technique (MM-DSSA), SSA decomposition is performed on the dissimilarity matrix of process data using MDS. In this method MDS handles the nonlinear correlation of the data in multiple modes in the decomposition procedure of basic SSA. That is, SSA is combined with MDS for nonlinear extraction of components in multiple modes. In multimodal KDSSA (MM-KDSSA), SSA is used to extract the multiple features using kernel multidimensional scaling (kernel MDS). In this technique, kernel MDS is used in basic SSA decomposition to handle the nonlinearity in the data. The proposed MM-NLSSA involves nonlinear extraction of multiple components of process signals obtained by reconstructing covariance structure of points in a low-dimensional space. On the other hand, MM-DSSA and MM-KDSSA extract nonlinear multiple components by reconstructing dissimilarities between pairs of data points by distance in a reduced dimensional space using linear and Gaussian kernels.

The proposed multimodal monitoring strategy extends the suitability of nonlinear SSA to statistically monitor processes based on nonlinear, auto-correlated measurements. Additionally, the resulting nonlinear PCA model at multiple modal is more sensitive in detecting changes in a process. These ideas are to be illustrated by suitable industrial data as indicated in chapter1. The basic idea in the above proposed methodology is that uncorrelated, nonlinear, principal components are analysed using conventional linear

statistical analysis techniques at different modes, which is equivalent to nonlinear analysis of original data. Application of the proposed technique is demonstrated using simulated data and the Tennessee Eastman Challenge process in this section and the next chapter.

4.1 NONLINEAR SINGULAR SPECTRUM ANALYSIS (NLSSA)

The objective of nonlinear SSA (NLSSA) is to capture both linear and nonlinear relationships between process variables by decomposing the data into multiple modes using data adaptive basis functions. This objective is achieved by replacing PCA with an auto-associative neural network in the decomposition process of SSA. Hsieu and Wu (2002) developed the NLSSA method based on NLPCA with a circular node at the network bottleneck (NLPCA.cir) to study the anharmonic nature of tropical stratospheric wind and pointed out that the general configuration of the NLPCA.cir is capable of extracting both open and closed curve solutions.

In the NLSSA method, the first three steps of SSA are used to pre-filter the data as described in chapter 3 (section 3.1), then the principal components of the first few leading SSA modes

$$\mathbf{t}(t) = \left[\mathbf{t}_1(t),\ \mathbf{t}_2(t), \ldots, \mathbf{t}_p(t)\ \right],\ t = 1,2,\ldots,K \qquad \textbf{(4.1.1)}$$

are used as input to a feed-forward neural network mapping through a bottleneck to the output $\mathbf{t}'$ as shown in Figure 4.1.

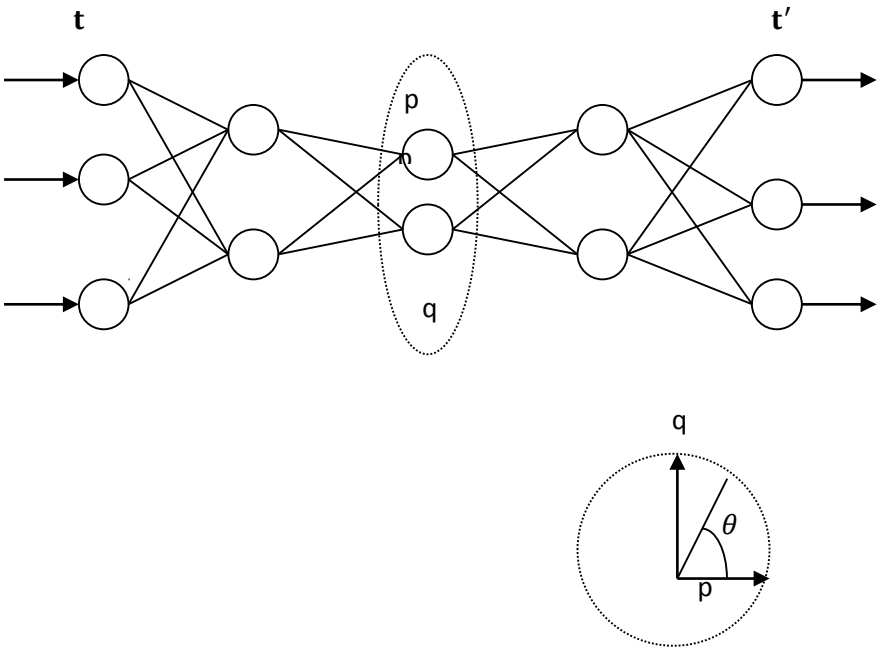


Figure 4.1: Schematic diagram of an auto-associative neural network (Hsieh, 2004) for approximating nonlinear principal components from SSA modes with a circular node at the bottleneck layer with two neurons p and q.

The model in Figure 4.1 is a standard, feed-forward neural network with three neurons denoted by circles enclosed between the input layer $\mathbf{t}$ on the left hand and the output layer $\mathbf{t}'$ on right-hand side. The encoding layer with $m = 2$ is placed next to the input layer with $l = 3$ neurons. The bottleneck layer with two neurons in the middle of the network is followed by a decoding layer with $m = 2$ neurons; and lastly, the output layer with $l = 3$ neurons.

Kirby and Miranda (1996) introduced the above network for extracting closed curve solutions. The original NLPCA network introduced by Kramer (1991) in Figure 2.4 is replaced with two neurons p and q in the bottleneck confined to lie on a unit circle with one degree of freedom by an angle θ as shown in Figure 4.1. The periodic or wave modes in the data can be extracted effectively using NLPCA.cir because of its ability to extract closed curve solutions (Hsieh, 2001).

In SSA, each periodic mode had to be split into a pair of modes such as stochastic and deterministic, since the analysis using PCA technique models the data by a straight line rather than by a closed curve (Hsieh & Wu, 2002). Thus, the use of NLPCA.cir can

combine two or more SSA modes into one NLSSA modes. The general configuration of NLPCA.cir models both open and closed curve solutions unlike the original NLPCA, which models only open curve solutions. Moreover, NLPCA.cir is more general than NLPCA and hence, is used to perform NLSSA (Hsieh & Wu, 2002).

As shown in equation (2.2.20) in chapter 2 (section 2.2.2), in NLPCA.cir the input vector $\mathbf{t}$ of length $K = N - M + 1$ is mapped to encoding layer (first hidden layer) $\boldsymbol{h}^{(\mathbf{t})}$ with elements

$$\boldsymbol{h}_k^{(\mathbf{t})} = \tanh\left(\left(\boldsymbol{w}^{(\mathbf{t})}\mathbf{t} + \mathbf{b}^{(\mathbf{t})}\right)_k\right),\ k = 1,2,\dots,m \qquad \textbf{(4.1.2)}$$

of length m and $\boldsymbol{w}^{(\mathbf{t})}$ is a weighted matrix with m rows and K columns, $\boldsymbol{b}^{(\mathbf{t})}$ contains the bias parameter of length m. Hyperbolic tangent functions are used as a transfer function $\boldsymbol{G}_1$ as in equation (2.2.20).

The mapping from the encoding layer to the bottleneck layer with coupled neurons p and q is performed with a linear transfer function. The initial states p_0 and q_0 are computed as

$$p_0 = \boldsymbol{w}^{(\mathbf{t})}.\boldsymbol{h}^{(\mathbf{t})} + \bar{b}^{(\mathbf{t})} \text{ and } q_0 = \widetilde{\mathbf{w}}^{(\mathbf{t})}.\boldsymbol{h}^{(\mathbf{t})} + \tilde{b}^{(\mathbf{t})} \qquad \textbf{(4.1.3)}$$

where $\boldsymbol{w}^{(\mathbf{t})}$, $\widetilde{\boldsymbol{w}}^{(\mathbf{t})}$ are weighted parameters and $\bar{b}^{(\mathbf{t})}$, $\tilde{b}^{(\mathbf{t})}$ are bias parameters.

The circular node is define as

$$p = {}^{p_0}/_{r} \text{ and } \ q = {}^{q_0}/_{r} \qquad \textbf{(4.1.4)}$$

where $r = (p_0^2 + q_0^2)^{\frac{1}{2}}$ and satisfies the unit circle equation $p^2 + q^2$=1 (Kirby & Miranda, 1996; Hsieh & Wu, 2002). Hence there is only one angular degree of freedom from the θ at the bottleneck layer because of the above circle constraint.

A hyperbolic tangent function is used for the mapping from the bottleneck layer to the decoding layer ($\boldsymbol{F}_1$), as shown in equation (2.2.23), which is given as

$$\boldsymbol{h}_k^{(\mathbf{s})} = \tanh((\boldsymbol{w}^{(\mathbf{s})}p + \widetilde{\boldsymbol{w}}^{(\mathbf{s})}q + \boldsymbol{b}^{(\mathbf{s})})_k),\ k = 1,2,\dots,m. \qquad \textbf{(4.1.5)}$$

The network output is represented by

$$\mathbf{t}_i' = (\boldsymbol{w}^{(\mathbf{s})}\boldsymbol{h}^{(\mathbf{s})} + \overline{\boldsymbol{b}}^{(\mathbf{s})})_i. \quad \textbf{(4.1.6)}$$

The optimal values of weight and bias parameters are determined by minimising the mean square error *m.s.e* of $\mathbf{t}'$ relative to $\mathbf{t}$, which is given as

$$J = \langle \|\mathbf{t} - \mathbf{t}'\|^2 \rangle. \quad \textbf{(4.1.7)}$$

For a better nonlinear solution, the number of hidden neurons $m \geq 2$ is used in the encoding and the decoding layers where as one hidden neuron is used in the bottleneck layer as shown in Figure 4.1. Hence, for any given number of hidden neurons in the input and output layer, one hidden neuron in the bottleneck layer will be enough to approximate any continuous function to arbitrary accuracy. The selection of number of neurons m in the encoding and decoding layer follow a general assumption that the nonlinear capability of the network is increased with a larger number of , but also give over-fitted solutions (Hsieh & Wu, 2002). The number of neurons (m) in the encoding and decoding layers is set to be equal in this study. Several rules have been used in the literature for the selection of number of neurons to avoid over-fitting in the training phase of the network. To have better generalisation performance, a procedure called early stopping was used in this study in which the *m.s.e* was monitored by gradually increasing the number of neurons and the training is stopped when *m.s.e* decreases steadily. Hence, the number of neurons which realise the minimum *m.s.e* is selected to avoid over-fitting in the training process.

The neural network outputs $\mathbf{t}'$ are the NLSSA mode 1 approximation of the first leading SSA modes (PCs). The NLSSA reconstructed component 1 (NLRC1) is the reconstruction of the original time series. The elementary matrix is obtained by multiplying

$$\mathbf{t}' = \left[\mathbf{t}_1', \ \mathbf{t}_2', \ldots, \mathbf{t}_p'\right] \quad \textbf{(4.1.8)}$$

by the corresponding SSA eigenvectors,

$$\mathbf{a} = \left[a_1, a_2, \ldots, a_p\right] \quad \textbf{(4.1.9)}$$

which are transformed into NLRC1 through diagonal averaging as explained in the step 4 of SSA decomposition.

The first nonlinear SSA (NLSSA) mode extracted from NLPCA.cir is a continuous curve solution, which is obtained by principal components nonlinearly. The residual vector obtained by subtracting the first NLSSA mode $\mathbf{t}'$ from $\mathbf{t}$ is the input to the NLPCA.cir network for the second NLSSA mode. This procedure can be repeated for subsequent modes.

4.2 MULTIMODAL NONLINEAR SSA (MM-NLSSA)

The MM-NLSSA is the PCA modelling of each reconstructed signal resulting from the multimodal decomposition of the original signal using NLSSA. The MM-NLSSA methodology consists of decomposing each variable by NLSSA into multiple modes, after which the PCA model is developed using the approximated variable at each mode. This accounts for the nonlinear correlation between the variables. Control limits for scores and residual at each mode are computed as in classical multivariate statistical process monitoring method (Kresta *et al*., 1991) using data representing normal operation. For new data, a fault or a significant change in the process parameter detects whether the scores or residuals of the reconstructed data violate the control limits at any mode. The proposed MM-NLSSA algorithm follows the monitoring framework shown in Figure 3.3 with use of NLSSA for decomposition of data into different modes. The MM-NLSSA also uses the same procedure for the selection of parameters in SSA decomposition in the linear case of multimodal SSA discussed in the previous chapter (section 3.2.1). The schematic summary of multimodal decomposition of NLSSA for the respective modes is illustrated in Figure 4.2.

The first step of MM-NLSSA, is the decomposition of the data matrix $\mathbf{X}(n \times m)$ with m variables and n observations, using NLSSA as described in the above section. That is each column of $\mathbf{X}$ is decomposed by applying SSA using a common window size M as discussed in the decomposition step of basic SSA in chapter 3 (section 3.1.1 and 3.1.2). In this study, the embedding window dimension is selected based on the first maximal decorrelation point of the autocorrelation function of each variable obtained by using equation (3.2.1). Thus, through SVD of the covariance matrix of the trajectory matrix $\mathbf{X}_j', j = 1,2,\dots,m$, the scores and the corresponding loading vectors are computed as

solution of the eigenvalue problems, as discussed in the SSA decomposition step. After SSA decomposition, the score matrix $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_m]$, where $\boldsymbol{T}_j = \mathbf{t}_{j,i}$, $j = 1,2, \dots, m$, $i = 1,2, \dots, p$ $j = 1, 2 \dots m$, $i = 1, 2, \dots p$ is computed by using equation (3.1.7) with the first p leading PCs, where each score $\mathbf{t}_{i,j}$ is a time series of length $K = n - M + 1$. A nonlinear PCA model is constructed for using each $\mathbf{T}_j$ as the inputs to the NLPCA.cir. The NLPCA reconstructed component 1, $\widetilde{\mathbf{T}} = \left[\widetilde{\mathbf{T}}_1 , \widetilde{\mathbf{T}}_2 , \dots, \widetilde{\mathbf{T}}_m\right]$ (NLPC1) is the approximation of $\mathbf{T}$ by the NLPCA mode 1. The trajectory matrix $\mathbf{X}_j'$ can be expressed in terms of an ordered series of nonlinear scores and loading vectors by multiplying these approximated nonlinear PCs (NLPCs) from the mode 1 with the corresponding SSA eigenvectors from SSA modes. That is

$$\mathbf{X}_j' = \widetilde{\mathbf{T}}_j \mathbf{P}_j^T = \tilde{\mathbf{t}}_{j,1}\mathbf{p}_{j,1}^T + \tilde{\mathbf{t}}_{j,2}\mathbf{p}_{j,2}^T + \cdots + \tilde{\mathbf{t}}_{j,p}\mathbf{p}_{j,p}^T, j = 1,2, \dots m. \qquad \textbf{(4.2.1)}$$

Summing over (diagonal averaging) p modes gives the reconstruction of the time series $\widetilde{\mathbf{X}}_j'$, $j = 1,2, \dots, m$, from the NLSSA mode 1, which can be considered as the multimodal approximation of mode 1. After the first NLSSA mode has been obtained, the nonlinear scores extracted from the first NLPCA.cir mode can be subtracted from the original scores to get the residuals. These residuals then are the input into the network to extract the second NLSSA mode. All other different modes are approximations of time series from respective NLSSA modes. Figure 4.2 illustrates the approximation of the i^{th}scale in MM-NLSSA.

Finally, as indicated in the monitoring procedure of MS-SSA and Figure 3.3, PCA is applied to the reconstructed data matrix $\widetilde{\mathbf{X}} = \left[\widetilde{\mathbf{X}}_1', \widetilde{\mathbf{X}}_2', \dots, \widetilde{\mathbf{X}}_m'\right]$ obtained by nonlinear linear transformation of the original data $\mathbf{X}$. The number of PCs and the control limits on the monitored statistics $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ at each mode are calculated as discussed in section (3.2.1.3). More specifically, each reconstructed matrix $\widetilde{\mathbf{X}}_j, j = 1,2, \dots, m$ in each mode is decomposed by PCA using equation (3.2.5). By using data representing normal operating conditions, the control limits for $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics are calculated. The $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ for sample k are computed using equations (3.2.6) and (3.2.7). The significance level at each mode is adjusted using equation (3.2.9).

For new data set $\mathbf{X}_{new}$ the lagged trajectory matrix $\mathbf{X}'_{new,j}$,$j = 1,2,\dots m$ for each variable is calculated using the window length M.The new score matrix is:

$$\mathbf{T}_{new} = \left[\mathbf{T}_{new,1}, \mathbf{T}_{new,2}, \dots, \mathbf{T}_{new,m}\right] \qquad \textbf{(4.2.2)}$$

where $\mathbf{T}_{new,j} = \mathbf{t}_{new,j,i}$, $j = 1,2,\dots,m$, $i = 1,2,\dots,p$ is obtained through eigenvalue decomposition of the resulting trajectory matrices and the first p leading PCs are used as the inputs to the NLPCA.cir. After extracting the first NLSSA mode $\widetilde{\mathbf{T}}_{new}$, the residual $\mathbf{T}_{new} - \widetilde{\mathbf{T}}_{new}$ can be the input to the second NLPCA mode and so on for all other higher modes.

The multimodal approximation of new data for each mode is obtained through the diagonal averaging of the reconstructed trajectory matrix $\widetilde{\mathbf{X}}'_{new,j}$, represented by the product of the new approximated nonlinear score matrix $\widetilde{\mathbf{T}}_{new,j}$ and the loading vector $\mathbf{P}_j$, where $j = 1,2,\dots,m$ obtained from the corresponding modes of the data operating under normal condition. The $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics for new data are computed using new scores obtained by the application of PCA on each reconstructed mode of the new data set using equations (2.2.12) and (2.2.13) and are compared with corresponding control limits calculated using normal data. The process is monitored to be out of control, if $\boldsymbol{T}^2$ or $\boldsymbol{Q}$ statistics of the reconstructed new data set exceeds the control limits at a specific mode.

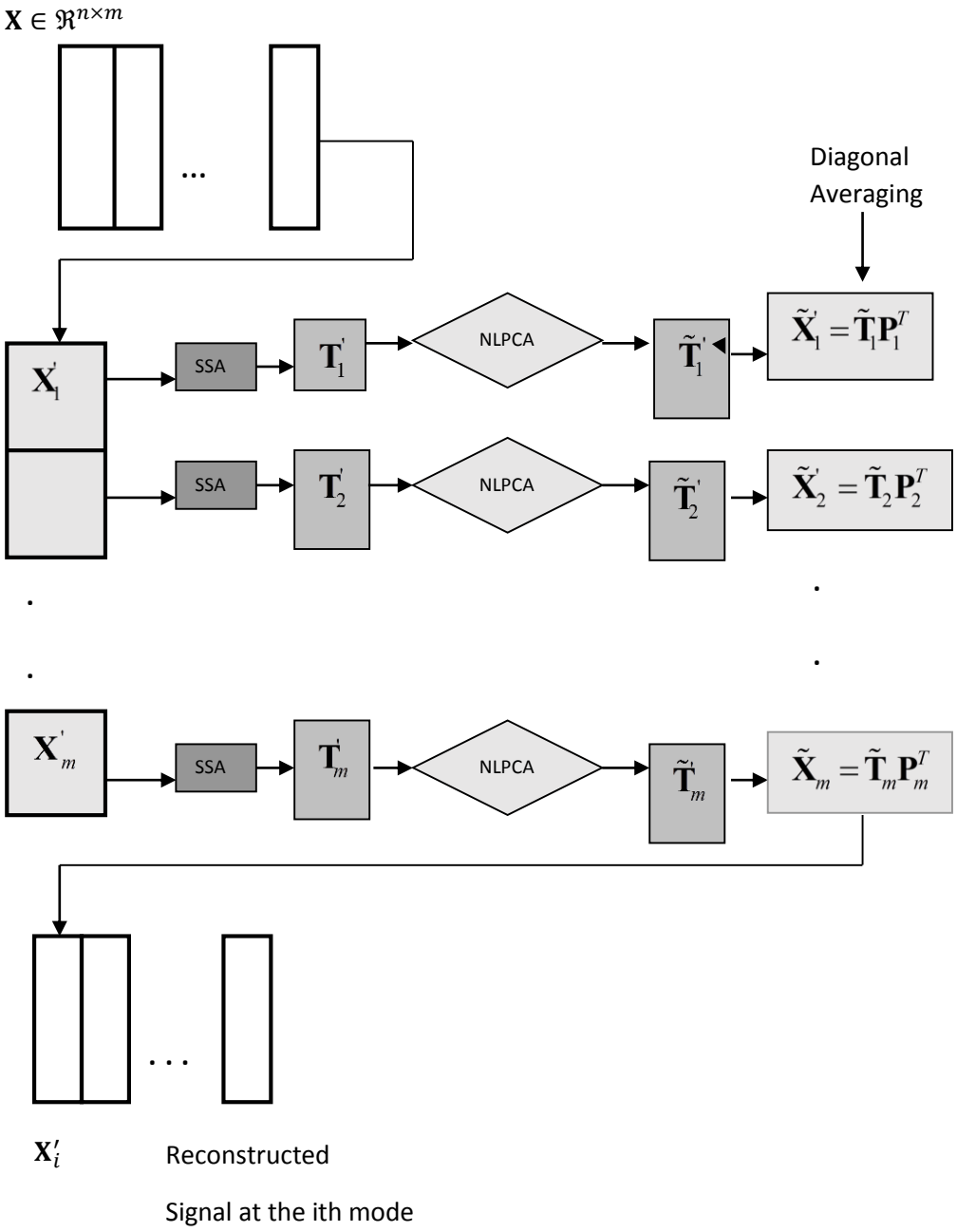


Figure 4.2: Schematic summary of multivariate NLSSA based multimodal decomposition.

MM-NLSSA methodology: An illustration

The MM-NLSSA algorithm is illustrated by data representing a simple mathematical example (Dong & McAvoy, 1996). The following nonlinear mathematical model consists of three variables x_1, x_2 and x_3 with added independent and identically distributed random noises e_1, e_2, and e_3.

$$x_1 = t + e_1 \quad \textbf{(4.2.3)}$$

$$x_2 = t^2 - 3t + e_2 \quad \textbf{(4.2.4)}$$

$$x_3 = -t^3 + 3t^2 + e_3 \quad \textbf{(4.2.5)}$$

where $e_i \in N(0, 0.01)$, $t \in [6, 15]$,is uniformly distributed and i = 1,2 ,3.

The first 900 samples generated by the above equations are taken as normal data. A fault condition (Case 1) is generated by changing the parameter of the

variable x_3 according to

$$x_3 = - 1.1t^{3.3} + 3.2t^{1.7} + e_3 \quad \textbf{(4.2.6)}$$

and keeping the other variables unchanged. A test data (Case 0) in which no fault was present was also generated as a control set for performance evaluation of MM-NLSSA. The data for normal and abnormal conditions are shown in Figure 4.3.

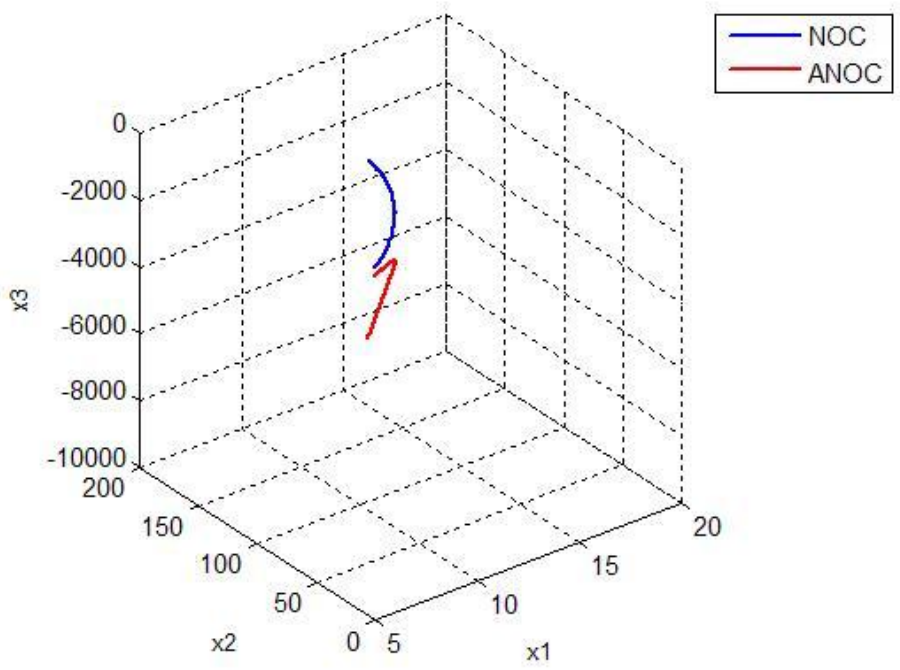


Figure 4.3: Data for normal operating condition (NOC) and abnormal operating condition (ANOC).

SSA with window size M = 331, is determined using equation (3.2.1) on the basis of the maximum value of the decorrelation point where the auto-correlation functions of the variables are zeros, as indicated in Figure 4.4. Here the size 331, corresponding to the first point of maximal decorrelation point of the three variables, was applied to the normal data. The first three PCs from each variable obtained by SSA decomposition accounted for at least 90% variance of the augmented data. These PCs ($\mathbf{T}_i$, $i = 1,2,\ldots,m$) are used as the input to NLPCA.cir network, with two hidden neurons in the encoding and decoding layers of the first variable, six hidden neurons for the second variable and three hidden neurons for the third variable respectively. The mean square error ($m.s.e$) between NLPCA.cir network output ($\widetilde{\mathbf{T}}_i$) and the original PC ($\mathbf{T}_i$) was minimum with the respective hidden neurons in those variables when compared with fewer or a larger selection of neurons in the hidden layers. Thus, the selection of the proper number of neurons in the encoding and decoding layer was essential to avoid the over fitting of the data for any false alarm in the monitoring chart.

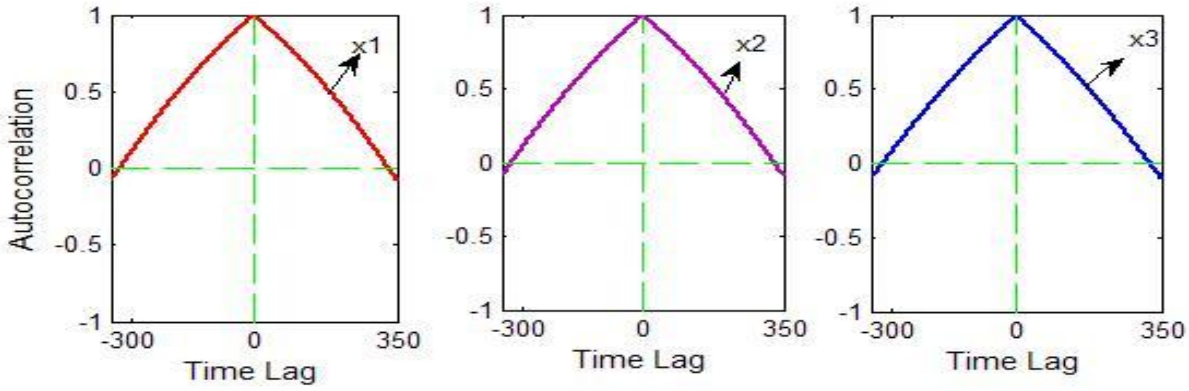


Figure 4.4: Sample autocorrelation functions of the variables. The first maximal decorrelation point is 331.

The NLPCA.cir network output for each variable ($\widetilde{\mathbf{T}}_i$) are the NLSSA mode 1 approximation for the three leading PCs. The NLSSA reconstructed components of the time series in the first mode/level is the nonlinear approximation of the original time series by NLSSA mode 1. This can be determined by multiplying the approximated PCs from NLPCA.cir with the corresponding SSA eigenvectors and diagonal averaging the matrices $\widetilde{\mathbf{X}}_i' = \widetilde{\mathbf{T}}_i\mathbf{P}_i^T = \left[\widetilde{\mathbf{T}}_1\mathbf{P}_1^T, \widetilde{\mathbf{T}}_2\mathbf{P}_2^T, \dots, \widetilde{\mathbf{T}}_m\mathbf{P}_m^T\right], i = 1,2,\dots,m$ for each variable. The residual ($\widetilde{\mathbf{T}}_i' = \mathbf{T}_i - \widetilde{\mathbf{T}}_i\,,\ i = 1,2,\dots,3$) of the approximated PCs from NLSSA mode 1 can be used as the input to the NLPCA. cir network for the second NLSSA mode. Multiplying the approximated PCs from the NLSSA mode 2 with the corresponding SSA eigenvectors and diagonal averaging the respective matrices as in the mode 1 extraction, give the components of the time series in mode/level two of the original data. These procedures can be repeated for all other modes with the respective residuals of the approximated PCs from the corresponding NLPCA.cir mode. The approximated signals in NLSSA modes are the closed curve solution of the original time series obtained with NLPCA.cir by combining two or more basic SSA modes. Thus, most of the features in the time series can be extracted with first few NLSSA modes to avoid the over- fitting of data. The *m.s.e* was found to be a minimum in the reconstructed components in mode 1 and 2 compared to that in higher modes. Hence, the first two modes from the first two NLSSA modes were chosen for PCA monitoring for fault detection. Figure 4.5 and Figure 4.6 show the reconstructed components of the normal and abnormal data in the first two modes/levels from mode 1 and 2. In Figure 4.5 the reconstructed components are the deterministic components of the original signal for the

both normal and abnormal data. On the other hand, in Figure 4.6 the reconstructed component of normal data represents noisy components while the reconstructed component of the test data represents the deterministic component of the data. Hence the faults in the process can be detected by using the first two modes.

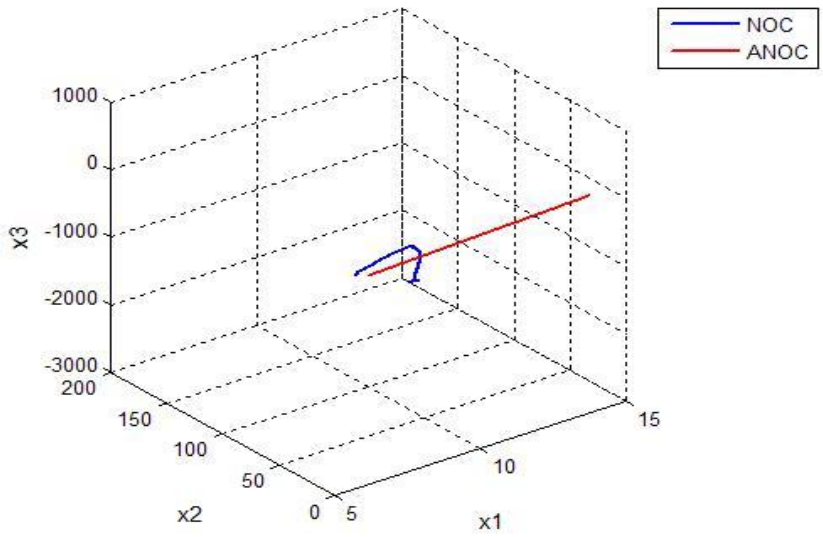


Figure 4.5: Reconstructed variables in normal operating conditions (NOC) and abnormal operating conditions (ANOC) in MM-NLSSA mode 1. These smothered reconstructed components represent the deterministic components of the data carrying more information.

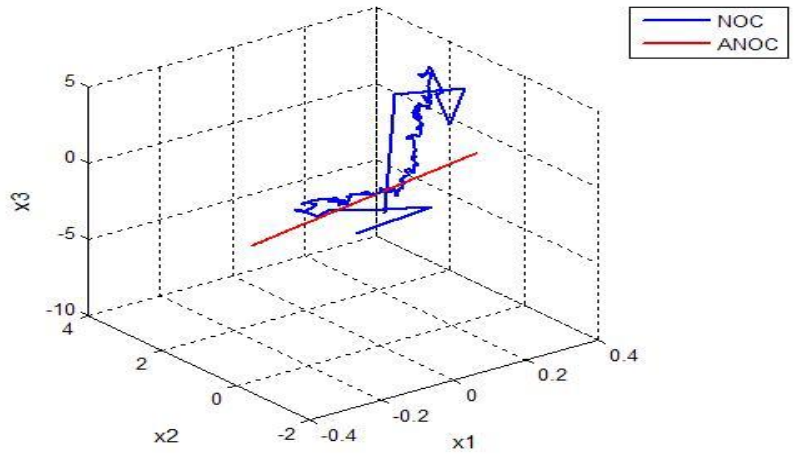


Figure 4.6: Reconstructed variables of normal and abnormal data in MM-NLSSA mode 2.The reconstructed variables in ANOC are smoother than the reconstructed variable in the NOC.The reconstructed variables in NOC are associated with noisy components.

The PCA model is used to monitor the reconstructed components in each mode and the SPE and $\boldsymbol{T}^2$ plots are used to detect the faults in each mode using equation (3.2.6) and equation (3.2.7). The control limits for both $\boldsymbol{T}^2$ and SPE statistics were set at 95% confidence level. The performance of MM-NLSSA in Case 0 and Case 1 are summarised in Table 4.1 on the basis of the percentage of reliability as shown in Figure 4.7 – 4.9. The reliability index is defined as the fraction of test samples violating 95% control limit in SPE and $\boldsymbol{T}^2$ plots. Any change in the process variables and the relationship among the process variables results in an increase in the $\boldsymbol{T}^2$ and SPE value of the process data, because the new measurements cannot be explained by the PCA model. Thus, any kind of abnormality in the process can be detected by monitoring both $\boldsymbol{T}^2$ and SPE charts of the continuous data.

Table 4.1: Reliability % of MM-NLSSA for Case 0 and Case 1

	Case 0		Case 1	
	T^2 statistics	**SPE statistics**	**T^2 statistics**	**SPE statistics**
Mode1	0%	1%	0%	83.9%
Mode 2	13%	16.8%	77.5%	32.9%

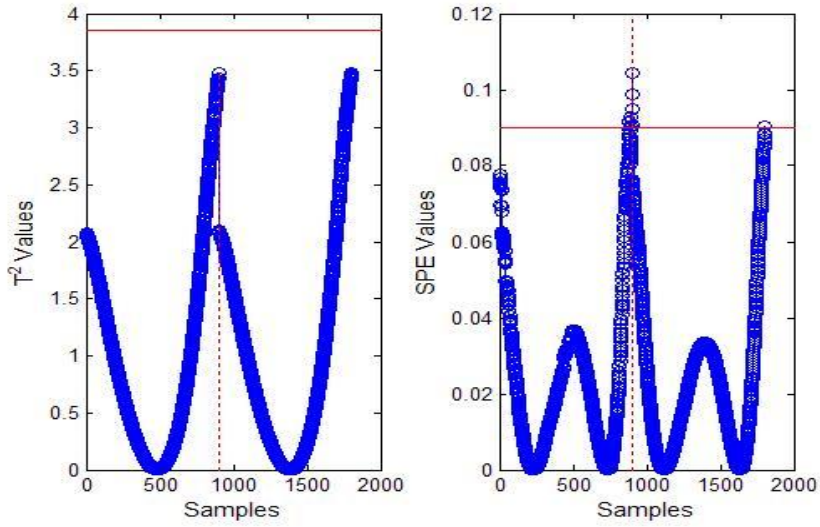


Figure 4.7: The T^2 and SPE chart for MM-NLSSA in Case 0 in mode1 for both the normal condition with the 95% confidence limit. The monitoring statistics are calculated using the first and last 900 samples taken from the normal condition of the simulated system. The dotted line separate the train and test data in the above figure.In T^2 chart all the samples are below the control limit while 1% test data samples exceeded the control limit in the SPE chart.

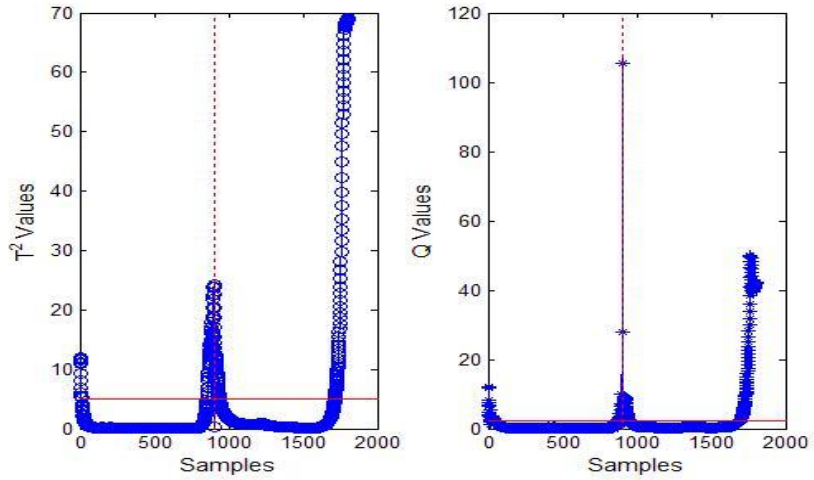


Figure 4.8: The T^2 and Q chart for MM-NLSSA in Case 0 in mode 2 with 95% confidence limit in the normal condition. In T^2 chart 13% of test samples in the normal condition violated the control limit while16.8% test samples exceeded the control limits in Q in mode 2. This variation in the reliability percentage from the expected percentage (5%) also indicates the presence of noise (high frequency components) in the reconstructed variables in mode 2 of the normal data in Case 0.

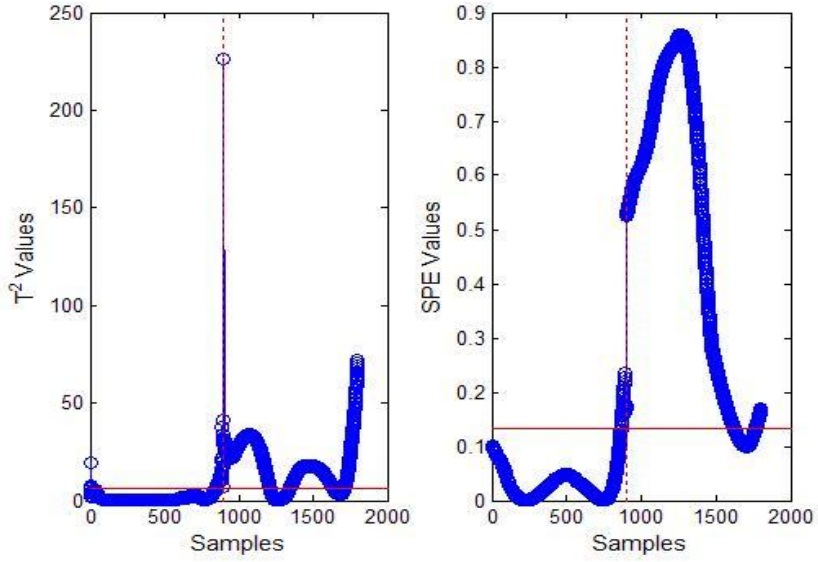


Figure 4.9: The T^2 and SPE chart for MM-NLSSA in Case1 for of both normal data data (first 900 samples) and faulty data (last 900 sample) with 95% confidence limit. The T^2 chart in mode 2 gave the highest reliability (77.5%) while in the SPE chart the highest relaibily percentage was (83.9%) obtained in mode 1.

The reliability measure obtained in the normal condition (Case 0) in the first mode is zero for $\boldsymbol{T}^2$ and 1% for the $\boldsymbol{Q}$ statistics. This indicates that calculations of control limits are successful in finding the deviations in the process when there are no faults. However, in the second mode the reliability percentage was above 5% due to the added noise in mode 2 of the reference model. The MM-NLSSA model detected the change in the variable 3(x_3) both in SPE and $\boldsymbol{T}^2$ chart as indicated in the table and in the Figure 4.9. The MM-NLSSA mode 2 detected the fault by extracting the nonlinear relationship among the PCs of SSA in both SPE and $\boldsymbol{T}^2$ chart while mode 1 detected the change in variable 3 only in the SPE chart. The deterministic reconstructed components in mode 1 in Figure 4.5 indicate that the components of the original signal in Figure 4.3 are approximated more accurately than that in Figure 4.6. Hence, a better detection of the parametric change of variable 3 is achieved in mode 1 of the $\boldsymbol{Q}$ chart. The parametric change in variable 3 causes the PCs in the test data to move away from the plane of the reference model. In addition, some points have overlap with the normal reconstructed data in the original plane of the reference model as shown in Figure 4.5. Thus, an increase in value of the SPE of the residuals is obtained compared to $\boldsymbol{T}^2$ values in mode 1. On the other hand, the $\boldsymbol{T}^2$ statistics is not very sensitive in mode 1 of MM-NLSSA since the reconstructed component of the faulty data within the model is not significant. This is the reason for the better performance of the SPE chart in mode 1. Nevertheless most of the reconstructed components in mode 2 of the test data are located within the model plane and some of them located outside the model plane as shown in Figure 4.6. Hence, $\boldsymbol{T}^2$ statistic detected the fault at higher number when compared to $\boldsymbol{Q}$ statistics. In general, the performance of MM-NLSSA in the first two modes shows that the proposed nonlinear SSA monitoring method can effectively detect the parametric changes in nonlinear process systems. More clearly, NLSSA effectively extracts nonlinear relations in PCs from different SSA modes while they are linearly uncorrelated.

Even though MM-NLSSA detected the faults in the process, this approach involves increased computational processing time due to the complexity in nonlinear optimisation of the NLPCA network for the extraction of nonlinear components in different modes. The creation of an optimum reference model based on 900 data samples also shows that the MM-NLSSA requires a massive amount of data for building the reference model as with the case of NLPCA models. In practical situations, to obtain a huge data set that operates in normal conditions is very difficult and is a challenge for the development of a

reference model. In addition, as discussed in the previous sections, the use of NLPCA in SSA decomposition by no means changes the nature of nonlinearity implication of the original multivariate data. A variation in SSA decomposition step is proposed in the next section in order to extract nonlinear components in multiple modes as an alternative solution for process monitoring to solve the issues with MM-NLSSA as mentioned earlier. In the proposed method the correlation between the variables are measured using the dissimilarity (inter-distance) scale structure in SSA decomposition instead of the variance-covariance (or correlation) matrix as in basic SSA. A multimodal process monitoring framework, as shown in Figure 3.3, that uses classical multidimensional scaling instead of PCA as in basic SSA decomposition, is proposed for monitoring processes that have nonlinear characteristics.

4.3 SINGULAR SPECTRUM ANALYSIS WITH DISSIMILARITY MATRIX (DSSA)

The objective of using a dissimilarity measures in SSA is to decompose time series data into multiple levels/modes of resolution, where the eigenvectors used for score projection are derived from a dissimilarity matrix instead of the correlation matrix used in basic SSA. The trajectory matrix $\mathbf{X}$ in equation (2.2.42) is scaled to zero mean and unit variance prior to the spectral decomposition step in SSA. Subsequently, the classical MDS algorithm is applied in the spectral decomposition of SSA. The multivariate scores are obtained through eigenvalue decomposition of double-centered squared dissimilarity matrix of size $K \times K$ using equation (2.2.40) on the scaled trajectory matrix $\mathbf{X}$ in terms of the number of observations in the process data. The squared dissimilarity matrix $\mathbf{D}^2$ on $\mathbf{X}$ is

$$\mathbf{D}^2 = \mathbf{X}\mathbf{X}^T. \tag{4.3.1}$$

The double centering equation in equation (4.3.2) below is then applied to $\mathbf{D}^2$ to obtain the inner product matrix $\mathbf{B}$;

$$\mathbf{B} = -\frac{1}{2}\,\mathbf{J}_K\mathbf{D}^2\,\mathbf{J}_K \tag{4.3.2}$$

where $\mathbf{J}_K = (\mathbf{I}_K - \mathbf{1}_K\,\mathbf{1}_K^T/K)$, $\mathbf{I}_K$ is an identity matrix, and $\mathbf{1}_k$ is a vector with elements of 1 and size K.

Using eigenvalue decomposition of $\mathbf{B}$ in equation (4.3.2) and using equations (2.2.41 - 44), the principal component scores can be computed as:

$$\mathbf{t}_k = \frac{1}{\sqrt{\lambda_k}} \mathbf{X}^T \mathbf{u}_k \tag{4.3.3}$$

where, $k = 1,2,\dots,K$, λ_k and $\mathbf{u}_k$ are the corresponding eigenvalues and eigenvectors of the double-centered dissimilarity matrix $\mathbf{B}$.

The dimension of the features or scores retained can be determined using the ratio of the retained and total eigenvalues (Cox & Cox, 1994), that is,

$$\sum_{i=1}^{p} \lambda_i / \sum_{i=1}^{K} \lambda_i \tag{4.3.4}$$

where λ_i is the i^{th} eigenvalue, p is the number of selected dimensions and $K = n - M + 1$ (i.e. number of samples in the dissimilarity matrix). Hence, for $p < K$, leading scores are selected based on equation (4.3.4) to represent the time series. The p-dimensional score vectors of the decomposed matrix $\mathbf{T}$ are given by

$$\mathbf{t}(t) = \left[\mathbf{t}_1(t), \mathbf{t}_2(t), \dots, \mathbf{t}_p(t)\right]^T, \; t = 1,2,\dots,K. \tag{4.3.5}$$

Hence the eigenvalue decomposition of the trajectory matrix can be written as:

$$\widetilde{\mathbf{X}} = \sum_{i=1}^{p} \tilde{\mathbf{x}}_i \tag{4.3.6}$$

where $\tilde{\mathbf{x}}_i = \sqrt{\lambda_i}\, \mathbf{u}_i \mathbf{t}_i^T$.

The multimodal approximations of the reconstructed signals in equation (4.3.6) can be achieved through diagonal averaging described in SSA decomposition (step 4).

4.4 MULTIMODAL NONLINEAR SSA WITH A DISSIMILARITY MATRIX (MM-DSSA)

The MS-DSSA methodology consists of decomposing each variable by DSSA into multiple modes after which the PCA model is developed using the approximated variable at each mode, thereby accounting for nonlinear correlation between the variables. Control limits for scores and residual at different modes are computed as in classical

multivariate statistical process monitoring method (Kresta *et al*., 1991) using data representing normal operation. For new data, a fault or a significant change in the process parameter is detected if the scores or residuals of the reconstructed data violate the control limits at any mode. The schematic summary of MM-DSSA procedure is shown in Figure 4.10 and the process monitoring procedure follows the basic process monitoring framework shown in Figure 3.3 where DSSA is used for decomposition step.

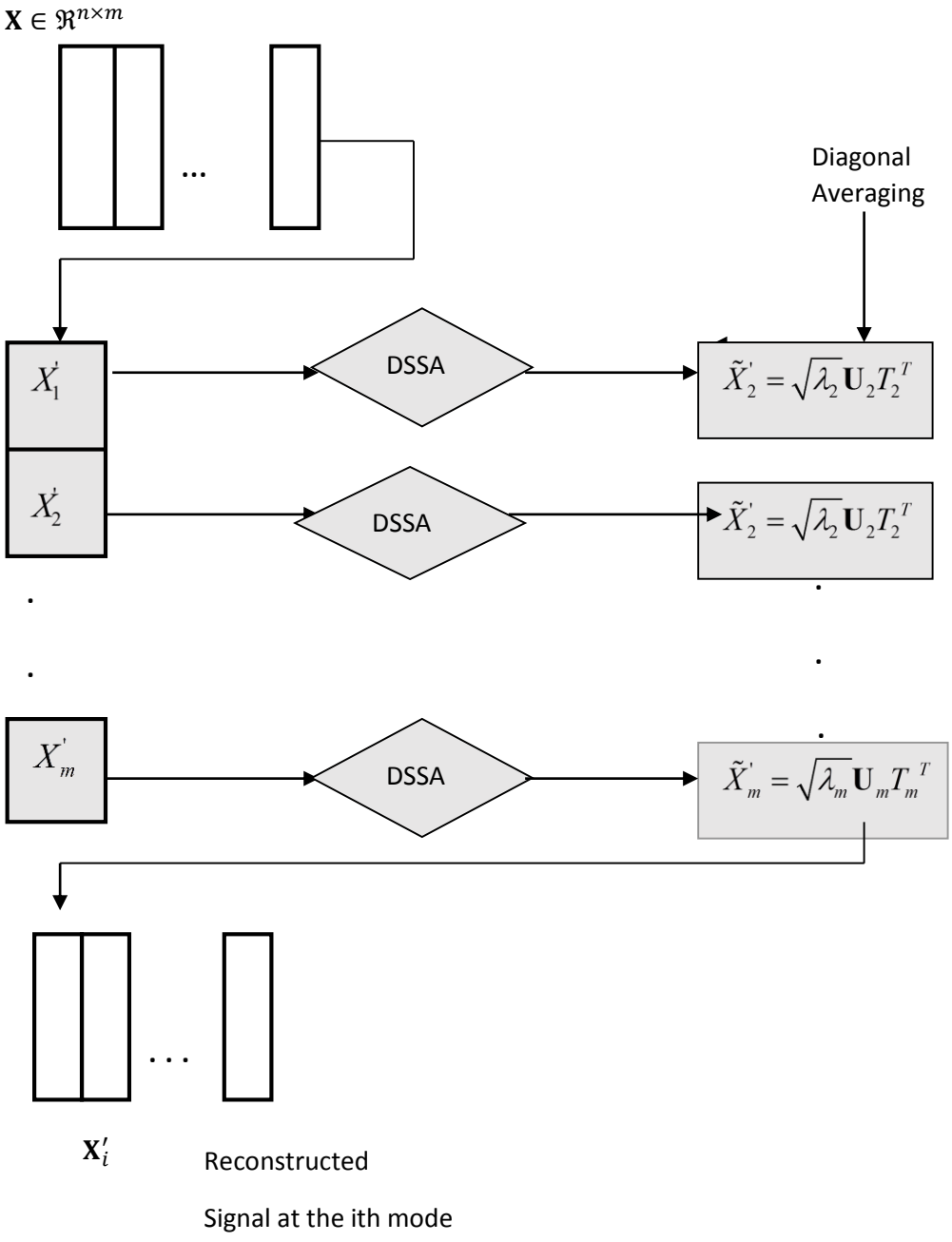


Figure 4.10: Schematic summary of decomposition of data using MS-DSSA at the ith scale.

The first step in MM-DSSA is the decomposition of the data matrix $\mathbf{X}(n \times m)$ with m variables and n samples by applying DSSA as explained in the previous section. More specifically, each column of $\mathbf{X}$ is decomposed by applying DSSA using a common window size M. The window length M is the first maximal decorrelation point of the variables in the process, which is obtained through equation (3.2.1). The scaled

trajectory matrix of each variable $\mathbf{x}_j, j = 1,2,\dots,m$ are expressed in lower dimension p ($p < K$) by decomposing them using the inter-distance scale/dissimilarity measures of sample points in order to obtain the scores and the corresponding loading vectors, that is:

$$\mathbf{X}'_j = \sqrt{\lambda_{j,1}}\,\mathbf{u}_{j,1}\mathbf{t}_{j,1}^T + \sqrt{\lambda_{j,2}}\,\mathbf{u}_{j,2}\mathbf{t}_{j,2}^T + \cdots + \sqrt{\lambda_{j,p}}\mathbf{u}_{j,p}\mathbf{t}_{j,p}^T. \qquad \textbf{(4.4.1)}$$

The original data matrix is then reconstructed by

$$\widetilde{\mathbf{X}}'_i = \left[r\left(\sqrt{\lambda_{1,i}}\mathbf{u}_{1,i}\mathbf{t}_{1,i}^T\right) \quad r\left(\sqrt{\lambda_{2,i}}\mathbf{u}_{2,i}\mathbf{t}_{2,i}^T\right) \quad \dots \quad r\left(\sqrt{\lambda_{m,i}}\mathbf{u}_{m,i}\mathbf{t}_{m,i}^T\right)\right] \qquad \textbf{(4.4.2)}$$

where $r(.)$ is the diagonal averaging function (equation (3.1.11)). Thus, the data matrix $\widetilde{\mathbf{X}}'_1$ in the first mode (level $i = 1$) will be represented by all m process variables reconstructed from their respective first eigenvector and score vector only.

$$\widetilde{\mathbf{X}}'_1 = \left[r\left(\sqrt{\lambda_{1,1}}\mathbf{u}_{1,1}\mathbf{t}_{1,1}^T\right) \quad r\left(\sqrt{\lambda_{2,1}}\,\mathbf{u}_{2,1}\mathbf{t}_{2,1}^T\right) \quad \dots \quad r\left(\sqrt{\lambda_{m,1}}\mathbf{u}_{m,1}\mathbf{t}_{m,1}^T\right)\right]. \qquad \textbf{(4.4.3)}$$

This is repeated for all p modes, resulting in p representations of the data. Lastly, analysis of the reconstructed data matrix $\widetilde{\mathbf{X}}' = \left[\widetilde{\mathbf{X}}'_1, \widetilde{\mathbf{X}}'_2, \dots, \widetilde{\mathbf{X}}'_p\right]$ of $\mathbf{X}$ obtained through equation (4.4.2) are performed using PCA as explained in the monitoring process of MS-SSA in chapter 3 (section 3.2.1). The number of PCs and the control limits on the monitored indices $\boldsymbol{T}^2$and $\boldsymbol{Q}$ at each mode are also determined accordingly.

That is each reconstructed matrix $\widetilde{\mathbf{X}}'_j$, $j = 1,2,\dots,p$ is decomposed using PCA and equation (3.2.5).The number of principal components at each mode and the control limits on $\boldsymbol{T}^2$and $\boldsymbol{Q}$ statistics are calculated using data operating at the normal condition. The $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics for sample k are obtained using equation (3.2.6) and equation (3.2.7).

For on-line monitoring, the lagged trajectory matrix of the new data set $\mathbf{X}_{new}$ is calculated using the window length M calculated in normal the operating condition. This is followed by calculating new scores of the new data set by projecting the lagged trajectory matrix onto the p loading vectors obtained in the normal operating condition. Thus, the new data set in different modes can be represented by all m process variables

reconstructed from their respective eigenvectors obtained in the normal operating condition and the new score vector as

$$\widetilde{\mathbf{X}}'_{new,i} = \left[r\left(\sqrt{\lambda_{1,i}}\mathbf{u}_{1,i}\mathbf{t}^T_{new,1,i}\right)\ r\left(\sqrt{\lambda_{2,i}}\mathbf{u}_{2,i}\mathbf{t}^T_{new,2,\mathrm{i}}\right)\ \ldots\ r\left(\sqrt{\lambda_{m,i}}\mathbf{u}_{m,\mathrm{i}}\mathbf{t}^T_{new,m,i}\right)\right] \quad \textbf{(4.4.4)}$$

where $r(.)$ is the diagonal averaging function.

The $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics for new data are calculated by using new scores obtained by using equation (3.2.8). The significance level at each mode is also adjusted using equation (3.2.9). If $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics for the reconstructed data at a specific mode exceed the control limit then the process is monitored to be out of control at that mode.

MM-DSSA methodology: An illustration

The MM-DSSA algorithm is evaluated by using the nonlinear data obtained through equations (4.2.3- 5) indicated in the previous section with three variables x_1, x_2 and x_3 with added independent and identically distributed random noises e_1, e_2, and e_3 (Dong & McAvoy, 1996), where $e_i \in N(0, 0.01)$, $t \in [10, 11]$ is uniformly distributed,and i = 1,2 ,3. The first 100 samples obtained by using equations (4.2.3-5) are taken as normal data. A fault condition (Case 1) is generated by changing the parameter of the variable x_3 according to equation (4.2.6) and keeping the other variables unchanged. A control set (Case 0) was also generated in which no fault was present for performance evaluation of MM-DSSA. The data for normal and abnormal conditions are shown in Figure 4.11.

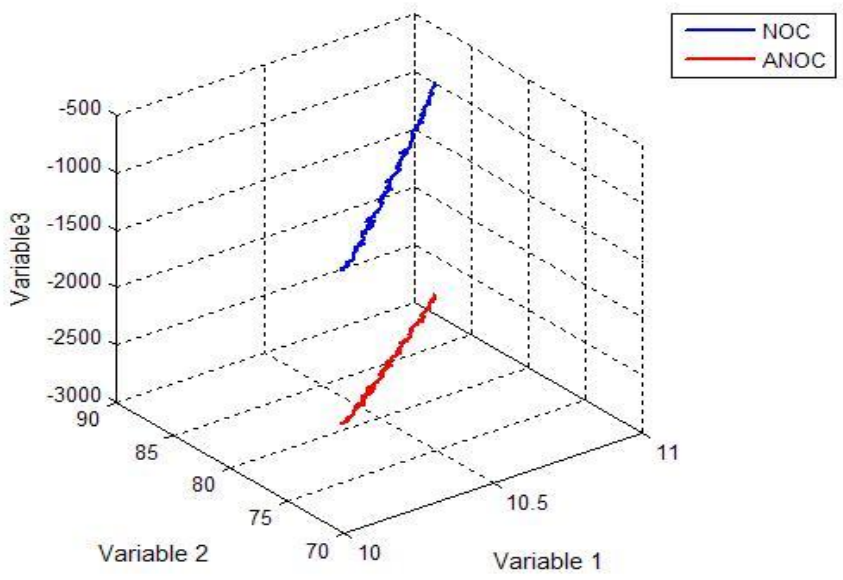


Figure 4.11: Data for normal operating condition (NOC) and abnormal operating condition (ANOC) in MM-DSSA.

An embedding window of size M= 37 was determined on the basis of the autocorrelation functions of the variables for SSA decomposition using equation (3.2.1) and is shown in Figure 4.12 as discussed in the previous chapter (section 3.2). The dissimilarity matrix on the trajectory matrices of each variable are computed using equation (4.3.1). The loading factors for each variable ($\mathbf{t}_i$, $i = 1,2,\dots,K, K = n - M + 1$)are obtained through the eigenvalue decomposition of double-centered dissimilarity matrix using equation (4.3.2) and equation (4.3.3). The selections of the score vectors to represent the original set of dissimilarities in multiple modes are determined based on equation (4.3.4) and also using the scree plot of the eigenvalues. The ratio of the retained and total eigenvalue is 95%.

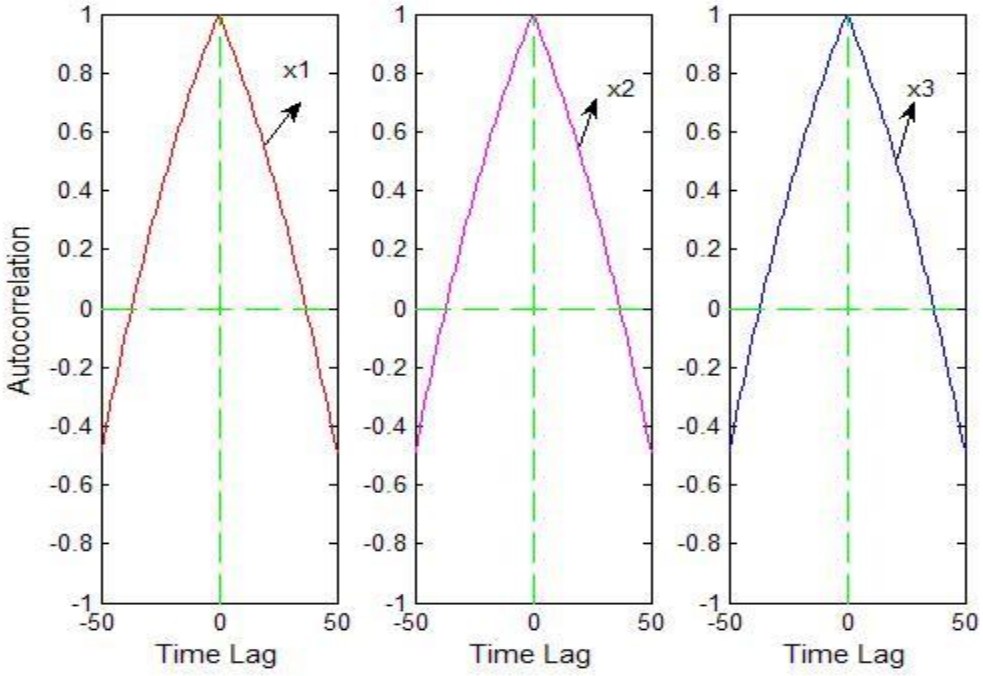


Figure 4.12: Sample autocorrelation functions of the variables. The first maximal decorrelation point is 37.

Figure 4.13 shows the plot of the logarithm of the singular values of the double centered dissimilarity matrices obtained from the lagged trajectory matrix of the normal data. The slowly decreasing sequence of eigenvalues can be produced by pure noise series while closely paired eigenvalues represent purely harmonic series (Golyandina *et al.*, 2001, Hassani, 2007, Hassani & Thomakos, 2010). Thus, a break or knee can also be used in the scree plot of the eigenvalues in Figure 4.13 to separate the different components in the signals such as trends and noise. A drop in the singular value occurs around component 12, which can be considered the start of the noise component. Hence, the first 12 principal scores from variables 1, 2 and 3 are selected to obtain the multimodal approximation of the original data based on the threshold from equation (4.3.4) as well as the scree plots in Figure 4.13.

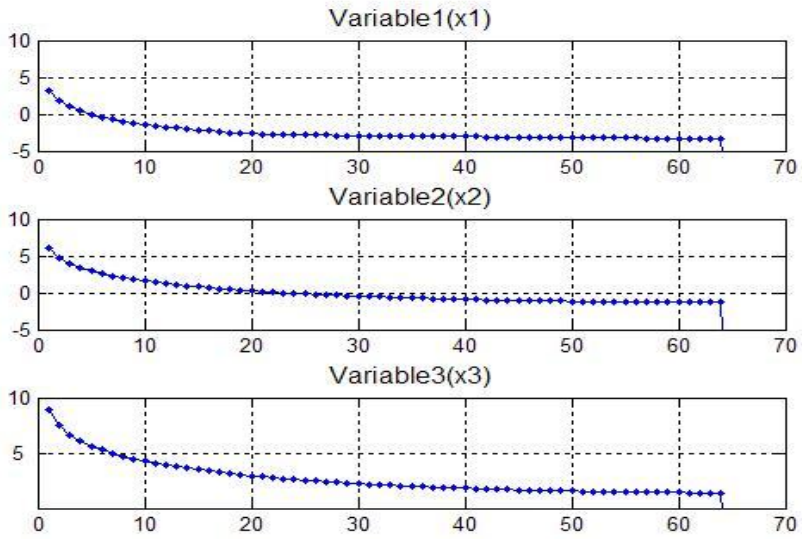
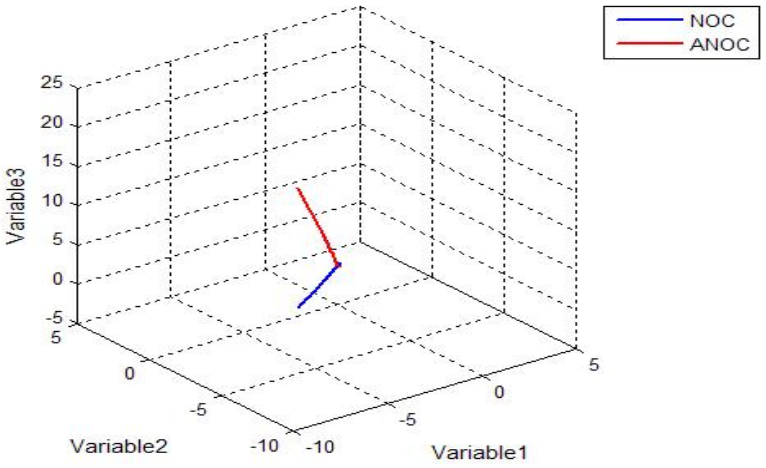


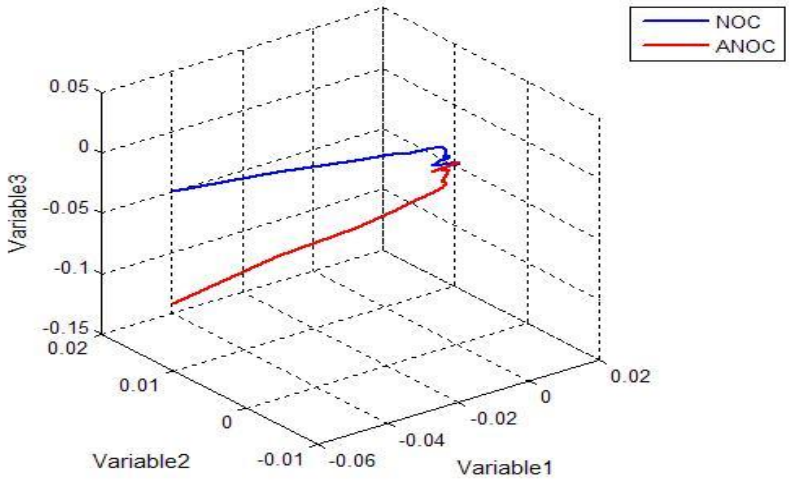
Figure 4.13: Logarithms of the eigenvalues of the dissimilarity matrix of the normal data. A drop in the eigenvalue occurs at 12th singular value.

The MM-DSSA reconstructed components of the original set of dissimilarities in mode 1 are approximated using equation (4.4.3), based on the first principal scores, eigenvalues and eigenvectors obtained from the double centered dissimilarity matrices of each variable. Thus, all other modes can be determined by multiplying the approximated scores with the corresponding eigenvectors and diagonal averaging the matrices for each variable.

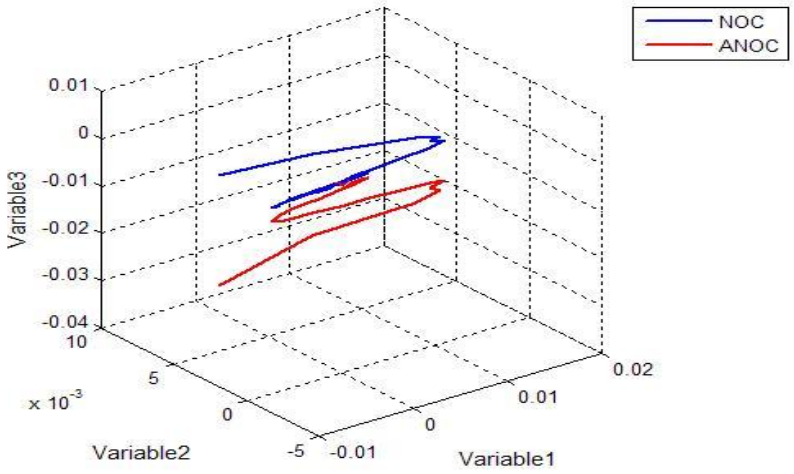
Hence, the first twelve modes from 65 modes were chosen for PCA monitoring for fault detection based on equation (4.3.4) and also based on the distribution of eigenvalues as shown in Figure 4.13. The selected modes represent different dynamical component of the original time series at different resolutions**.** The first twelve principal scores obtained from DSSA decomposition accounted for at least 95% variation of the augmented data from each variable. The percentage provides the sense of proximity of the reconstructed scorers with respect to the measures of dissimilarity scales (in terms of sample inter-distance measure) instead of variance (Yunus & Zhang, 2010). Figure 4.14 (a-d) shows the reconstructed components of the normal and abnormal data in the first, second fourth and the last mode (twelve).



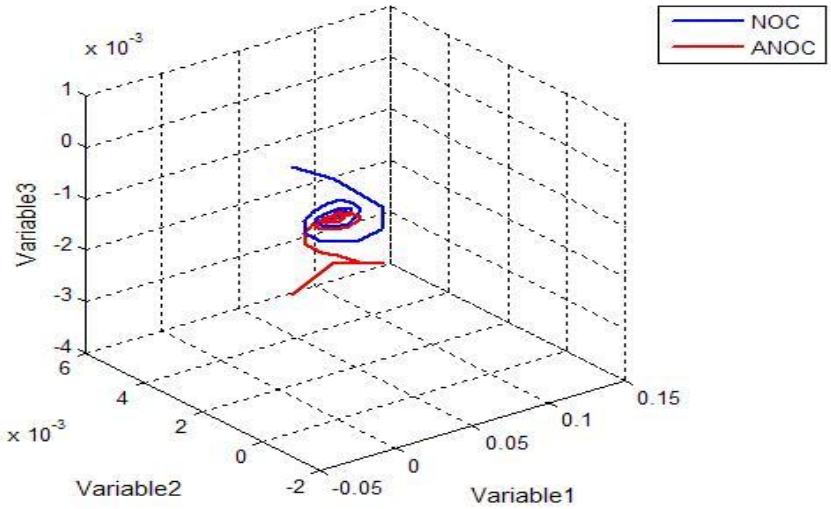
(a)



(b)



(c)



(d)

Figure 4.14 (a), (b), (c) and (d): Reconstructed variables in the normal operating condition (NOC) and abnormal operating condition (ANOC) in mode 1, 2, 4 and 12 for MM-DSSA. The smoother reconstructed variables in NOC and ANOC in mode 1, 2 and 4 in figure 23 a-c are associated with deterministic components. Figure 23 (d) shows the reconstructed variables in mode 12 which is associated with the noisy components.

The reconstructed components in each mode are monitored using PCA as explained in the monitoring step of MS-SSA in the previous chapter (section 3.2.1.3). $\boldsymbol{T}^2$ and SPE plots are used to detect faults in each mode using equation (3.2.6) and (3.2.7). The control limits for both statistics were set at 95% confidence level. The performance of

MM-DSSA is summarised in Table 4.2 based on the percentage of reliability as shown in Figure 4.15-4.16. The reliability index is defined as the fraction of test samples violating the 95% control limit in T^2 and SPE plots. Any kind of abnormality that is not explained by the PCA model of the reconstructed components results in an increase in the T^2 and SPE values of the process data as mentioned earlier. The reliability measure obtained in Case0 for MM-DSSA is close to the expected false alarm rate of 5% as shown in Figure 4.15. Therefore, the control limits are successfully determined in the normal operating condition to detect the faults in abnormal condition.

Table 4.2: Reliability % of MM-DSSA

Monitoring statistics	Case0	Case 1
T^2	8%	15%
Q	5%	99%

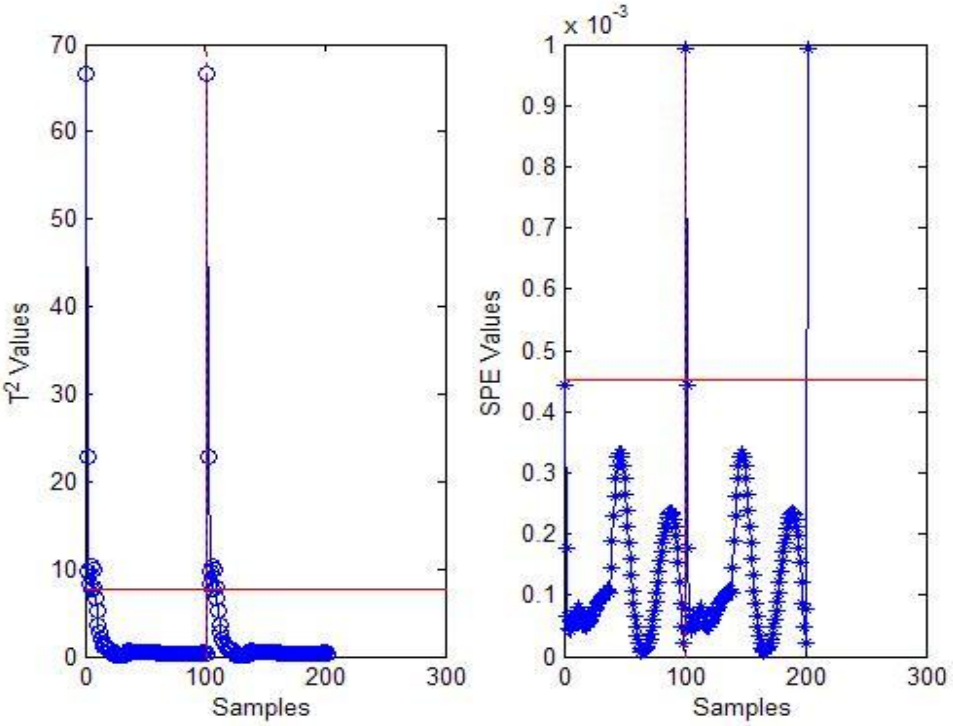


Figure 4.15: The T^2 and SPE chart for MM-DSSA in Case 0 with 95% confidence limit. The first 100 samples represent the normal train data and the last 100 samples represent the test data from normal operating condition. In these charts the number of samples exceeded the control limits are within the expected range.

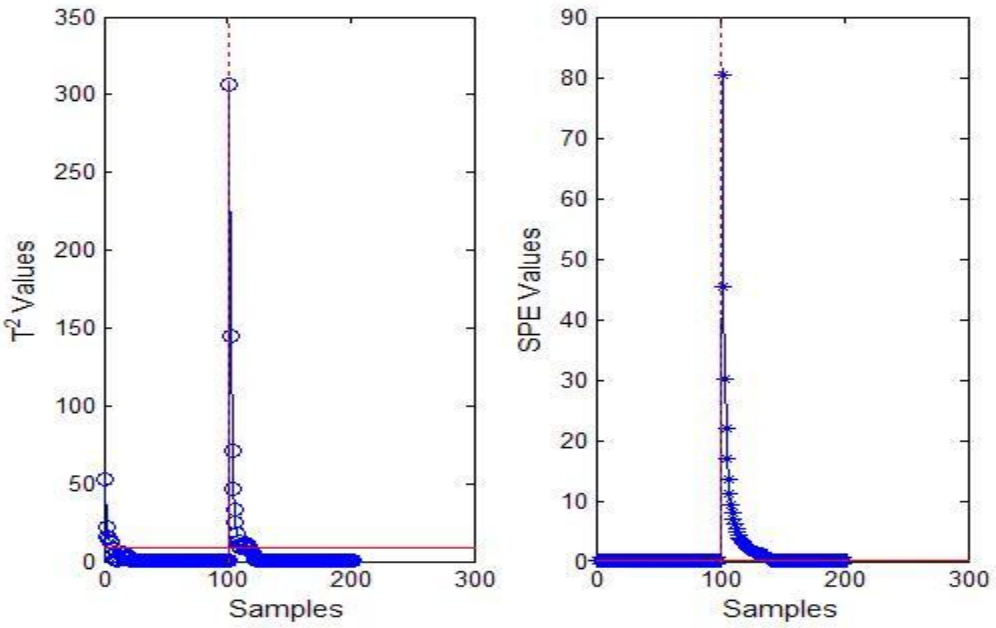


Figure 4.16: The T^2 and SPE chart for MM-DSSA in Case 1 with 95% confidence limit for both the normal data (first 100 samples) and the faulty data (last 100 samples). In the T^2 chart 15% (highest reliability) of samples in faulty data exceeded the control limits in mode 4. The highest reliability (99%) in the SPE chart is obtained from the reconstructed components in mode 1.

The MM-DSSA model detected the change in variable 3 (x_3) in the SPE chart of mode 1 as indicated in Table 4.2 and in the Figure 4.16. The better performance of MM-DSSA was detected at mode 4 and 1 out of 12 modes in the $\boldsymbol{T}^2$ and SPE chart respectively. The high reliability percentage obtained in the SPE chart in mode 1 in this study clearly shows that the parametric change in variable 3 (x_3) broke the correlation in the normal data that caused the variation in the residual space; hence, the reconstructed component of faulty data is placed outside the reference model plane as shown in Figure 4.14(a). Thus, MM-DSSA captured the nonlinear correlation in data in the first mode, which is the deterministic component of the data as shown in Figure 4.14(a) and leads to the reduced dimensionality selection to build the PCA model of the reconstructed signals. Hence, monitoring of reconstructed signals in those modes using PCA leads to a reduction in the computational time and cost of process monitoring task. In this context MM-DSSA can resolve the issues with conventional methods when monitoring a nonlinear data as well as the issue of nonlinear optimisation in NLSSA when using

NLPCA.cir. On the other hand, the issue of using a massive amount of data used to build the optimised reference model in the previous study can be resolved by using MM-DSSA in process monitoring.

The variation in SSA decomposition that was proposed in the above section used a dissimilarity measure to decompose time series data into multiple levels/modes of resolution in which the scores are projected on eigenvectors derived from a dissimilarity matrix instead of correlation matrix as used in basic SSA. This is equivalent to using CMDS procedure instead of PCA in the decomposition step of basic SSA to extract and project principal scores, as described in the previous section. The use of kernel functions in data analysis provides a nonlinear mapping of data and recently became a popular tool for data reduction in process monitoring as discussed in Chapter 2 (section 2.2.3). The duality between KPCA and CMDS discussed in kernel MDS in section 2.2.4 also shows that CMDS can be used for nonlinear dimensionality reduction by using kernels as similarities where the points in the Euclidean space are equal to the input vectors. Hence, the devised SSA framework based on CMDS extracts and approximates the nonlinear components in the data into multiple levels/modes (Zhang & Jordan, 2009). In this study we used Gaussian kernels for the nonlinear mapping from input space to the Euclidean space. This is also considered as using a kernel MDS algorithm (Williams, 2002) in SSA decomposition and named the *singular spectrum analysis with kernel MDS (KDSSA*), which is described in the following section.

4.5 SINGULAR SPECTRUM ANALYSIS WITH KERNEL MDS (KDSSA)

In KDSSA, the trajectory matrix $\mathbf{X}$ in equation (3.1.2) is scaled to zero mean and unit variance prior to the spectral decomposition step in SSA. Subsequently, the kernel MDS algorithm is applied in the spectral decomposition of SSA. The multivariate scores are obtained through eigenvalue decomposition of a double centered kernel matrix of size $K \times K$ using equation (2.2.40) and Gaussian kernels (equation (2.2.39)) on the scaled trajectory matrix $\mathbf{X}$ in terms of the number of observations in the process data. The squared dissimilarity matrix $\mathbf{D}^2$ on $\mathbf{X}$ is determined by equation (4.3.1).

The kernel matrix $\boldsymbol{K}$ is obtained through the squared exponential kernel. That is,

$$\boldsymbol{K} = \boldsymbol{k}(x_i, x_j) = exp\left\{-\theta(x_i - x_j)^T(x_i - x_j)\right\} = \exp(-\theta \mathbf{D}^2). \quad \textbf{(4.5.1)}$$

In this study the parameter θ is 0.5. Then the double-centering equation in equation (4.5.1) below is applied to $\boldsymbol{K}$ to obtain the inner product matrix $\mathbf{B}$;

$$\mathbf{B} = \mathbf{J}_K \boldsymbol{K} \mathbf{J}_K \quad \textbf{(4.5.2)}$$

where $\mathbf{J}_K = (\mathbf{I}_K - \mathbf{1}_K \mathbf{1}_K^T/K)$, $\mathbf{I}_K$ is an identity matrix, and $\mathbf{1}_K$ is a vector with elements of 1 and size K.

Eigenvale decomposition on $\mathbf{B}$ in equation (4.5.2) yields the reconstruction of the trajectory matrix

$$\widetilde{\mathbf{X}} = \sum_{i=1}^{p} \frac{1}{\sqrt{\lambda_i}} \mathbf{V}_i \mathbf{V}_i^T \boldsymbol{K} \quad \textbf{(4.5.3)}$$

where $\mathbf{V}_i$ is the eigenvector and λ_i is the corresponding eigenvalue of $\mathbf{B}$ (double centered kernel matrix) and $i = 1,2, \dots, p$.

The dimension of the retained eigenvectors can be determined using the ratio of the retained and total eigenvalues (Cox & Cox, 1994) as shown in equation (4.3.4). Hence, for $p < K$ leading scores are selected based on equation (4.3.4) to represent the time series.

The multimodal approximations of the reconstructed signals in equation (4.5.3) can be obtained through diagonal averaging described in SSA decomposition (step 4).

4.6 MULTIMODAL NONLINEAR SSA WITH KERNEL MDS (MM-KDSSA)

In MM-KDSSA, each variable is decomposed into multiple modes by KDSSA. The PCA model is then built using the reconstructed variable at each mode for the extraction of nonlinear correlation between the variables. The schematic summary of the MM-KDSSA algorithm is shown in Figure 4.17. The monitoring algorithm is developed based on the process monitoring framework provided in Figure 3.3, using KDSSA for the decomposition step. By using data representing a normal operation, control limits for scores and residuals at different modes are computed as shown in the classical

multivariate statistical process monitoring method (Kresta *et al.*, 1991). For new data, if the scores or residuals of the reconstructed data violate the control limits at any mode indicates the occurrence of a fault or a significant change in the process parameter.

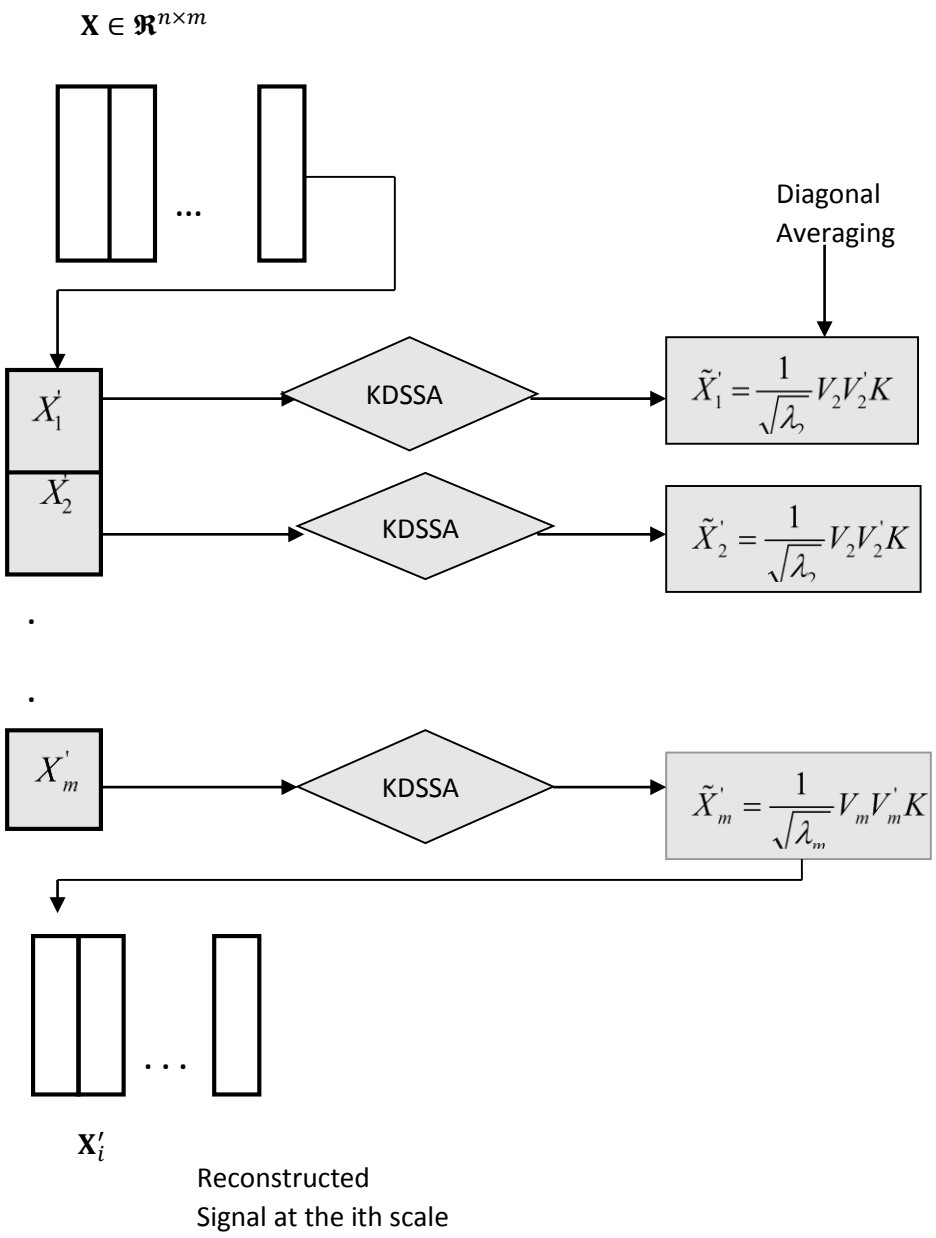


Figure 4.17: Schematic summary of decomposition of data using MM-KDSSA at the ith mode.

The first step in MS-KDSSA is the decomposition of each variable in the data matrix $\mathbf{X}\ (n \times m)$ with m variables and n samples by applying KDSSA using a common window size M as explained in section 4.5. The window length M for embedding the data in the delay coordinate space can be selected such that the points of different lagged vectors in each variable $\mathbf{X}_i(l)$, $\mathbf{X}_i(k)$, $(l \neq k), 1 \leq i \leq m$ are linearly independent. In this study, the

window length M is selected as the first maximal decorrelation point of variables in the process system using equation (3.2.1).

The scaled trajectory matrix of each variable $\mathbf{x}_j, j = 1,2,\dots,m$ is expressed in lower dimension p ($p < K$) by decomposing them using double-centered kernels as described in section 4.5 (equation (4.5.2)). This is obtained through Gaussian kernels in which the squared inter-distance scale/dissimilarity measures of sample points in each variable served as input space; that is

$$\mathbf{X}_j' = \frac{1}{\sqrt{\lambda_{j,1}}}\mathbf{v}_{j,1}\mathbf{v}_{j,1}^T\boldsymbol{K} + \frac{1}{\sqrt{\lambda_{j,2}}}\mathbf{v}_{j,2}\mathbf{v}_{j,2}^T\boldsymbol{K} + \cdots + \frac{1}{\sqrt{\lambda_{j,p}}}\mathbf{v}_{j,p}\mathbf{v}_{j,p}^T\boldsymbol{K}. \qquad \textbf{(4.6.1)}$$

The original data matrix can be reconstructed by

$$\widetilde{\mathbf{X}}_i' = \left[r\left(\frac{1}{\sqrt{\lambda_{1,i}}}\mathbf{v}_{1,i}\mathbf{v}_{1,i}^T\boldsymbol{K}\right) \quad r\left(\frac{1}{\sqrt{\lambda_{2,i}}}\mathbf{v}_{2,i}\mathbf{v}_{2,i}^T\boldsymbol{K}\right)\dots \quad r\left(\frac{1}{\sqrt{\lambda_{m,i}}}\mathbf{v}_{m,i}\mathbf{v}_{m,i}^T\boldsymbol{K}\right)\right] \qquad \textbf{(4.6.2)}$$

where $r(.)$ is the diagonal averaging function (equation (3.1.11)). Here the data matrix $\widetilde{\mathbf{X}}_1'$ in the first mode (level $i = 1$) represents all m reconstructed process variables from the corresponding first eigenvector and eigenvalues only.

$$\widetilde{\mathbf{X}}_1' = \left[r\left(\frac{1}{\sqrt{\lambda_{1,1}}}\mathbf{v}_{1,1}\mathbf{v}_{1,1}^T\boldsymbol{K}\right) \quad r\left(\frac{1}{\sqrt{\lambda_{2,1}}}\mathbf{v}_{2,1}\mathbf{v}_{2,1}^T\boldsymbol{K}\right)\dots \quad r\left(\frac{1}{\sqrt{\lambda_{m,1}}}\mathbf{v}_{m,1}\mathbf{v}_{m,1}^T\boldsymbol{K}\right)\right]. \qquad \textbf{(4.6.3)}$$

This procedure is repeated for all p modes, resulting in p multimodal representations of the data. Finally, approximation of the reconstructed data matrix $\widetilde{\mathbf{X}}'$ of , obtained through equation (4.6.2), are monitored using PCA as explained in section 3.2. That is, each reconstructed matrix $\widetilde{\mathbf{X}}_j'$ is decomposed using PCA, where $j = 1,2,\dots,p$ as explained in section (3.2.1.3) using equation (3.2.5). That is, the p approximations of the original data are subsequently monitored separately.

Similar to conventional PCA, the appropriate numbers of principal components retained at each scale are selected and the control limits on the monitored indexes (Hotelling's $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics) are determined using a data set obtained under normal operating conditions. The Hotelling's $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistic for sample k are calculated using equation (3.2.6) and equation (3.2.7).

For on-line monitoring, the lagged trajectory matrix of the new data $\mathbf{X}_{new}$ is computed using the common window M obtained in the normal condition. Then the kernel matrix $\boldsymbol{K}_{new} = \boldsymbol{k}(x_i, x_j)$ is computed using equation (4.5.1), where $x_i \in \mathbf{X}$ and $x_j \in \mathbf{X}_{new}$. The reconstruction of the trajectory matrix of the test data can be computed as

$$\widetilde{\mathbf{X}}'_{new} = \sum_{i=1}^{p} \frac{1}{\sqrt{\lambda_i}} \mathbf{V}_i \mathbf{V}_i^T \boldsymbol{K}_{new} \qquad \textbf{(4.6.4)}$$

where $\mathbf{V}_i$ is the eigenvector and λ_i is the corresponding eigenvalue of double-centered kernel matrix of $\mathbf{X}$. The multimodal approximation of the test data is obtained by performing the diagonal averaging of the above matrix (equation (4.6.4)) by equation (3.1.11). The values of $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics are also calculated for the new data sets, i.e. the scores of the new data are calculated by projecting the reconstructed test data onto the retained principal component loadings calculated by equation (3.2.8). The significance level at each scale are adjusted using equation (3.2.9). If, at a specific scale, $\boldsymbol{T}^2$or $\boldsymbol{Q}$ statistic for the reconstructed new data set is outside the calculated control limits, the process is judged to be out of control.

MM-KDSSA methodology: An illustration

The MM-DSSA methodology is illustrated by data representing a simple mathematical example obtained by using equation (4.2.3- 5), as discussed in section 4.2.1 and section 4.4.1.

The first 100 samples generated by the above equations are taken as normal data.

A fault condition (Case 1) is generated by changing the parameter of the variable x_3 according to

$$x_3 = -1.1t^3 + 3.2t^2 + e_3 \qquad \textbf{(4.6.5)}$$

and keeping the other variables unchanged. A test data (Case 0) in which no fault was present was also generated as a control set for performance evaluation of MM-KDSSA. The data for normal and abnormal conditions are shown in Figure 4.18.

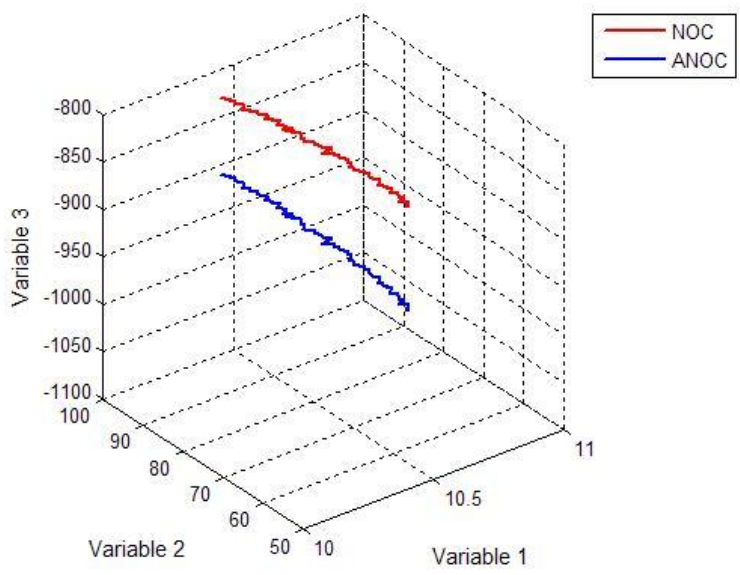


Figure 4.18: Data for normal operating condition (NOC) and abnormal operating condition (ANOC) in MM-KDSSA.

An embedding window of size M =37, for SSA decomposition using equation (3.2.1), was determined on the basis of the maximum value of the decorrelation point where the auto-correlation functions of the variables are zero, as discussed in chapter 3 (section 3.2.1) and shown in Figure 4.12. The dissimilarity measure of the trajectory matrices of each variable are computed using equation (4.3.1). The data points in the dissimilarity matrix are mapped into a feature space by using the Gaussian (exponential) kernel function in equation with θ = 0.5 as discussed in section 4.5 (equation (4.5.1)). The double centered kernel matrix $\mathbf{B}$ is computed using equation (4.5.2). The eigenvalue (λ_i) and the eigenvector ($\mathbf{V}_i$) for each variable ($i = 1,2 \ldots, K, K = n - M + 1$) are obtained through the eigenvalue decomposition of double-centered kernel matrix in equation (4.5.2) for the reconstruction of the trajectory matrix using equation (4.5.3). The selection of the eigenvalues and eigenvectors to represent the original set of dissimilarities in multiple modes are determined based on equation (4.3.4) and also using the distribution of the eigenvalues. The ratio of the retained and total eigenvalue explains at least 95% of variance in this study. Figure 4.19 shows the plot of the logarithm of singular values of the double centered kernel matrix of the dissimilarity measures calculated from the lagged trajectory matrices of the data under normal condition. A plateau in the eigenspectra shows the components with almost equal singular values. In this study the

first three eigenvalues are selected, because a drop in singular value occurred around the third component in Figure 4.19. Hence, the first three eigenvalues and eigenvectors from variables 1, 2 and 3 are selected to obtain the multimodal approximation of the original data based on the threshold from equation (4.3.4) as well as the scree plots in Figure 4.19.

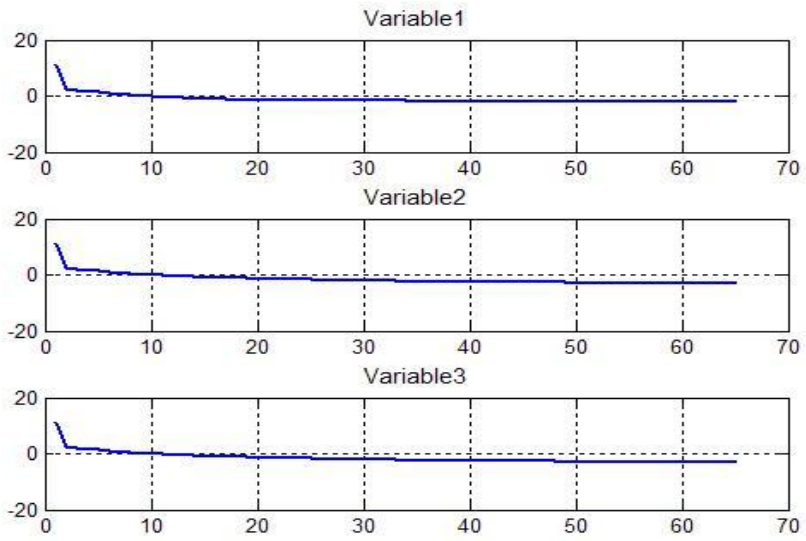
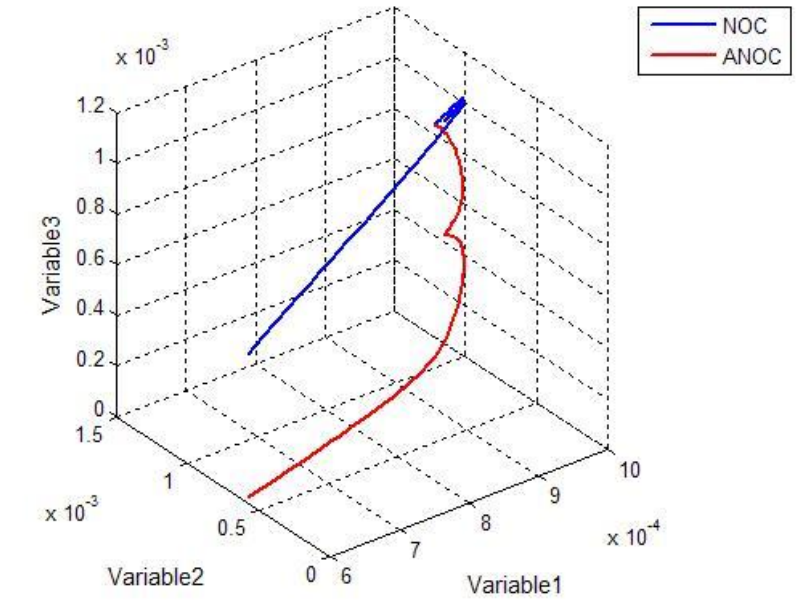
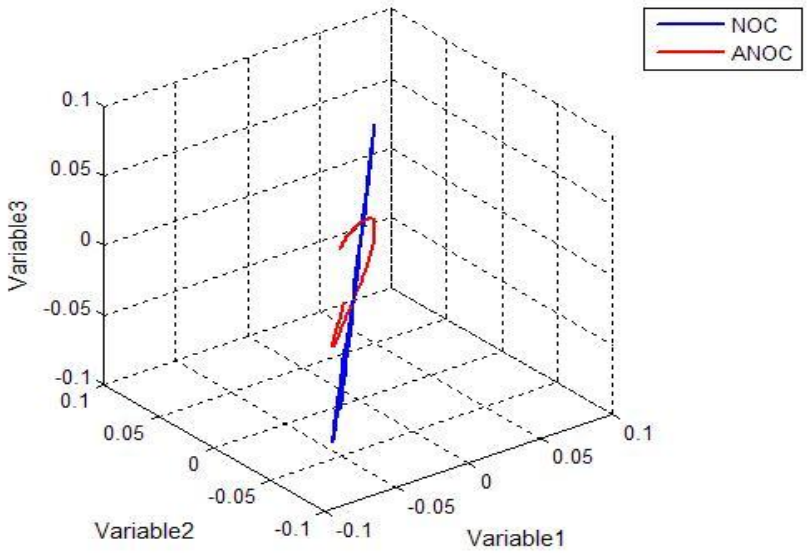


Figure 4.19: Logarithms of the eigenvalues of the kernel dissimilarity matrix of the normal data. A drop in the eigenvalue occurs at the 3rd singular value.

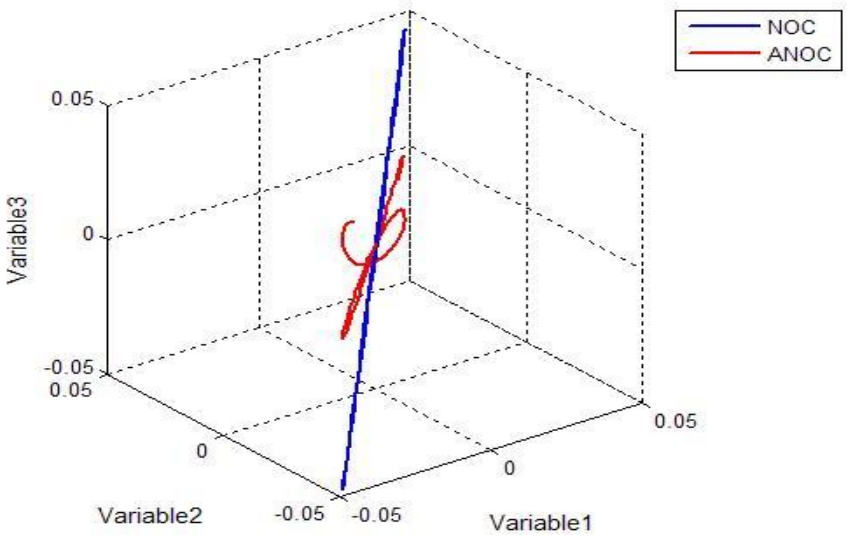
The variables are reconstructed into multiple levels through diagonal averaging of the data obtained by projecting the transformed dissimilarity measure onto the normalised eigenvectors using equation (4.6.1) and (4.6.2). The extracted components at different levels/modes represent different characteristics of the time series. The first three modes from 65 modes are selected for PCA monitoring that is explained in section (3.2.1.3) for fault detection. Figure 4.20(a-c) shows the reconstructed components of the normal and abnormal data in the first three modes. In Figure 4.20(a-c) the reconstructed components are the deterministic components of the original signal. Hence, most of the information in the signal can be extracted just by monitoring the deterministic component of the data in those modes.



(a)



(b)



(c)

Figure 4.20: Reconstructed variables in the normal operating condition (NOC) and the abnormal operating condition (ANOC) in mode 1, 2, and 3 for MM-KDSSA. The smoother reconstructed variables in the normal and abnomal data in the first three modes represent deterministic components.

The first three eigenvectors obtained from the KDSSA decomposition accounted for at least 99% variation of the augmented data from each variable. The $\boldsymbol{T}^2$and $\boldsymbol{Q}$ statistics plots are used to detect faults in each mode using equation (3.2.6) and (3.2.7). The control limits for both statistics were set at 95% confidence level.

The performance of MM-KDSSA is summarised in Table 4.3 based on the percentage reliability as obtained from Figure 4.21.The reliability index is defined as the percentage of test samples violating the 95% control limit in $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ plots as discussed in the previous studies. Any kind of change that is not captured by the PCA reference model of the reconstructed components results in an increase in the $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ value of the process data as mentioned earlier. The reliability measure obtained in Case 0 for MM-KDSSA is also close to the expected false alarm rate of 5% as shown in Figure 4.22. Thus, the control limits are successfully computed in the normal operating condition to detect the faults in abnormal situation.

Table 4.3: Reliability % of MM-KDSSA

Method	Monitoring Statistics	Case0	Case1
MM-KDSSA	T^2	8%	39%
	Q	5%	100%

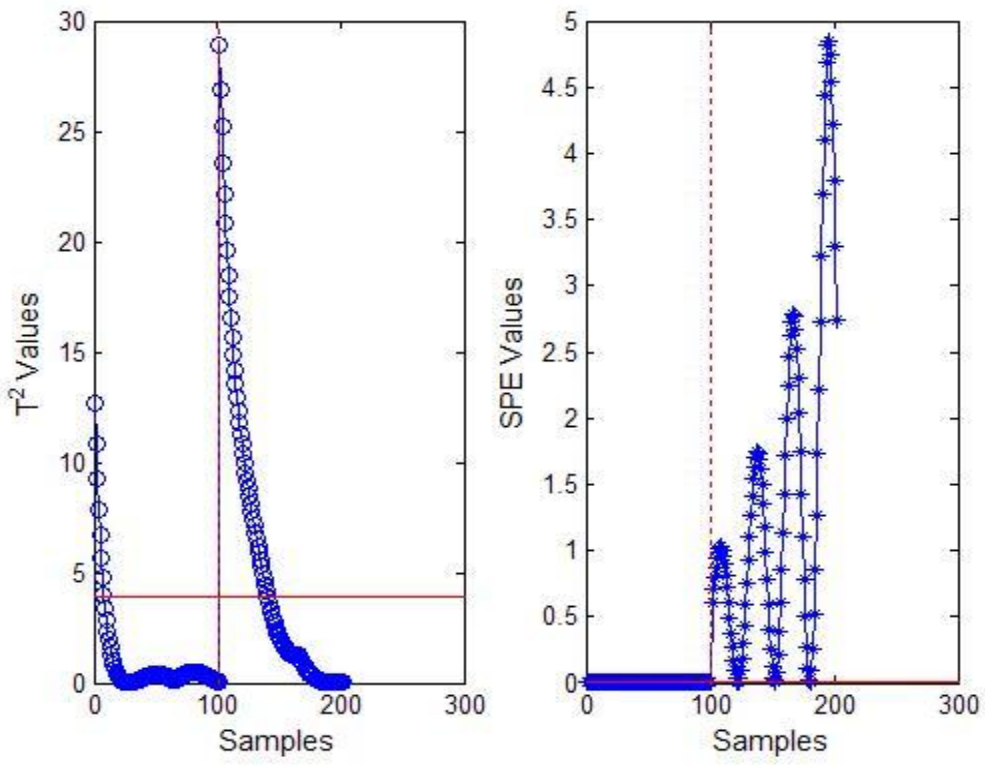


Figure 4.21: The T^2 and SPE chart for MM- KDSSA in Case 1 with 95% confidence limit for both nomal data (first 100 samples) and faulty data (last 100 samples). The reconstructed components in the first mode gave the highest reliability (39%) in T^2 chart while the components in the 3rd mode gave 100% reliability in SPE chart.

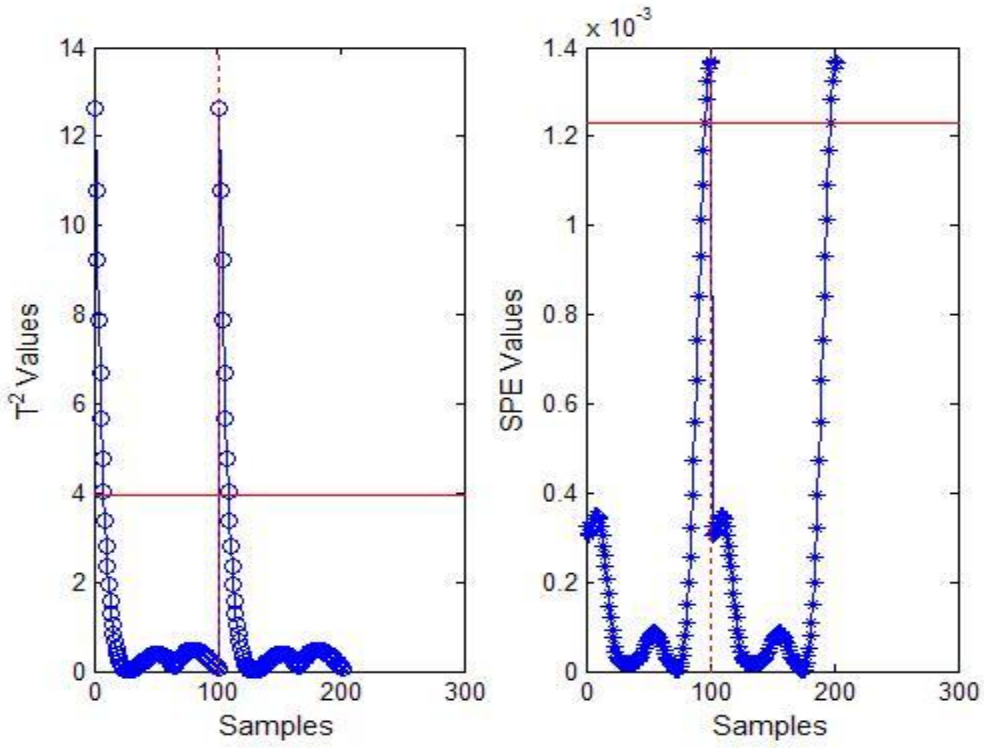


Figure 4.22: The T^2 and SPE chart for MM- KDSSA in Case 0 with 95% confidence limit for the normal condition. The first 100 samples were taken as train data and the last 100 samples were taken as test data.The number of samples exceeded the control limit in these charts were within the expected reliability percentage.

The value of $\boldsymbol{Q}$ statistics exceeded the confidence limit in all the three modes and the highest reliability percentage were obtained in mode 3, as indicated in the Table 4.3 and in Figure 4.21. This shows that $\boldsymbol{Q}$ statistics captured the systematic variation due to the parametric change in the variable (x_3) better than $\boldsymbol{T}^2$ statistics. On the other hand, $\boldsymbol{T}^2$ statistics are not very sensitive to the fault in this case. This is because the variability of the faulty data within the model is not significant and the parametric change in variable three caused a large difference between the variables and its projection onto the latent variables in the model .This is the reason for the better performance of $\boldsymbol{Q}$ chart as shown in the Figure 4.20(a) and 20(c). The results in Table 4.3 also indicate that performance of MM-KDSSA in detecting the change in variable x_3 was successful and best detection for both statistics occurred at the first mode and the third mode. This clearly indicates that MM-KDSSA, benefiting from the linear data structure using Gaussian kernels in the feature space, could capture the nonlinearity in the input space

to obtain the better or comparable monitoring results in this study. Hence, MM-KDSSA can also resolve the issues with conventional methods when monitoring nonlinear data with a dissimilarity structure in basic SSA decomposition.

In conclusion, the process parameters are physical process coefficients, which are directly related to the monitored process. Thus, the deviations of the process parameter from their normal value can be realated to a significant change in the process.The proposed nonlinear SSA methods showed better performace in detecting changes in the process parameter in the above illustrated studies with nonlinear SSA; hence, better dection of faults in the process system.

CHAPTER 5: MULTIMODAL NONLINEAR SSA: PROCESS MONITORING AND COMPARISON

In this chapter, the proposed methodologies of MM-NLSSA, MM-DSSA and MM-KDSSA are evaluated and compared by means of three case studies: (i) the Tennessee Eastman Challenge problem used as a benchmark in many process control studies (Downs & Vogel, 1993), and (ii) industrial process from Base Metal Flotation Plant. The performance of the proposed multimodal nonlinear methods are compared with the classical multivariate process monitoring method based on PCA as well as process monitoring based on multimodal basic SSA (MS-SSA). Dissimilarity plots are used to visualise the global structure of the similarity between different sample points in different positions relative to each other in Tennessee Eastman Challenge process. These plots are used to understand the relationship between sample points in three difficult fault cases in the Tennessee Eastman process (TE-Process), which are unobservable from the process data.

5.1 CASE STUDY 1: BASE METAL FLOTATION PLANT

In this study, MM-KDSSA, MM-DSSA and MM-NLSSA monitoring methods are applied to industrial data from an Australian base metal flotation plant. In this context, the proposed methods can be used for early detection of any abnormal event that could potentially lead to sudden changes in the plant. These changes mainly occur due to the disturbances in the grinding circuit as well as to the zinc rougher concentrate grade, which are directly related to the textural features/structure of the froth in the zinc roughers. This is achieved by identifying variations in textural features or parameters of the froth such as froth mobility, froth stability, bubble size, colour and distribution, which were characterised by the neighbouring grey level dependence matrix method. The quantification of these features is a difficult task and any error in the control of the process can lead to serious problems in process operation. Thus, the proper monitoring and control of the variations in the image features of the froth are useful in tracing the

cause or origin of those changes in troubleshooting efforts. Changes in the froth appearance could predict changes in metullargical performance, such as instabilities in the grinding circuit owing to mechanical failures and random changes in the attributes of the ore (Bezuidenhout *et al*., 1997, Aldrich *et al*., 2000).

In industrial processes, prior assumptions with regard to process operations are difficult to obtain unlike in other case studies discussed in this study. This is because the quality of data is unknown and data is possibly corrupted with unknown faults and gross errors. Moreover, the dependencies of the froth features and plant performance on ore characteristics are nonlinearly related. Hence, the application of multimodal nonlinear methods to monitor the froth textures will perform better than classical linear MSPC methods to separate deterministic and stochastic components in data for an improved safer process operation.

In this study, five features (image variables) are extracted from digital images of surface froths in the zinc roughers using the neighbouring grey level dependence matrix method as discussed in the work by Bezuidenhout *et al*. (1997). These variables are called small number emphasis feature (SNE), second moment (SM), average grey level of the image (AGL), entropy feature(ENT), and the instability of the froth (INSTAB). The flotation plant concentrator consists of a few integrated unit operations, namely, Grinding mill (SAG), Ball mill, Hydrocyclone for classification of pulp from ball mill, flotation feed buffer tank and rougher banks. The camera is placed to monitor the froth surface of the third zinc rougher. A simplified layout of the concentrator of the base metal flotation plant is shown in Figure 5.1.

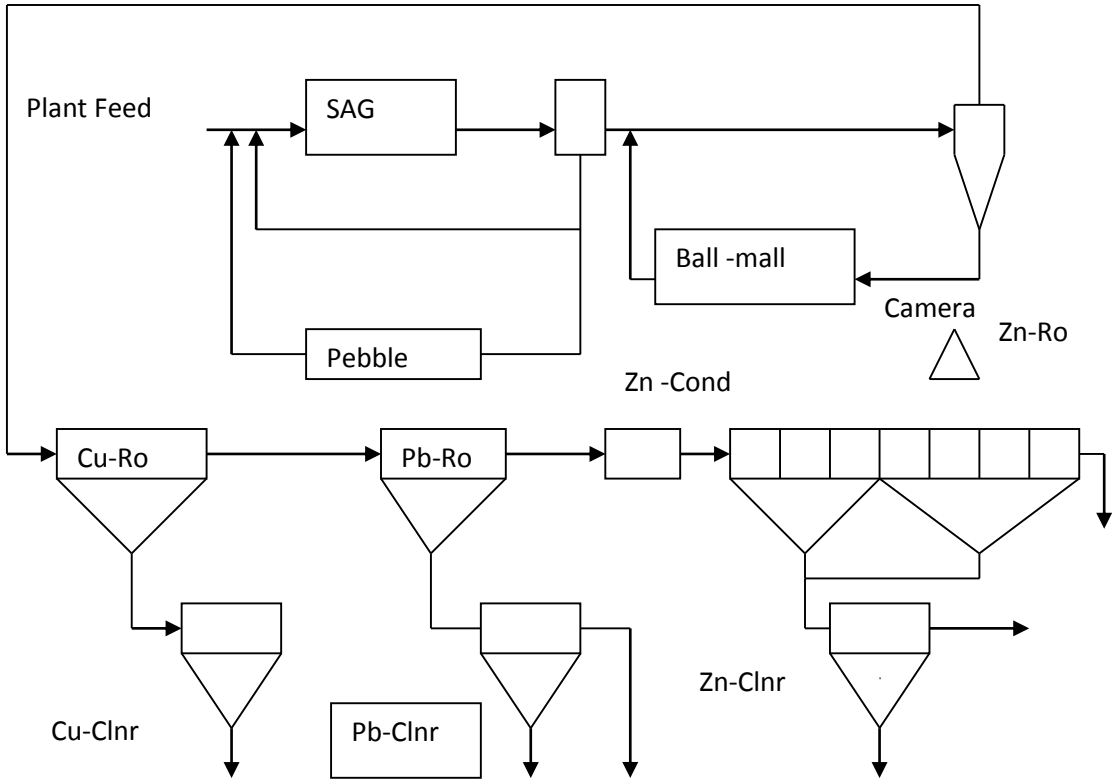


Figure 5.1: The schematic diagram of the concentrator of a base metal flotation plant (Bezuidenhout *et al*., 1997).

The MM-KDSSA, MM-DSSA, MM-NLSSA and MS-SSA based process monitoring algorithm was applied using an embedding window of size M = 43, determined on the basis of the auto-correlation functions of the variables using equation (3.2.1). The reference model for PCA, MS-SSA, MM-DSSA, MM-KDSSA were developed based on a set of 284 samples collected under normal operating conditions. A set of 284 samples collected under faulty conditions was projected onto each of the models for evaluation. The faulty data was collected during the change in froth image features because of the variations in the operating condition due to various reasons such as instability in grinding circuits, random variations in ore feed, stoppage in SAG mill and variations in the dosage of addition of reagents in the zinc rougher bank. The variables representing normal and abnormal conditions are displayed in Figure 5.2. First three modes from 242 modes are monitored for fault detection in MM-KDSSA, while 43 modes in MS-SSA and 206 modes from 242 modes were monitored in MM-DSSA and the first two modes were monitored in MS-NLSSA. The percentages of variance explained by selected eigenvalues for reconstruction of components in those three modes were at least 95% in

MS-KDSSA, which was obtained by equation (4.3.4) and by the spectrum of the logarithm of eigenvalues. In MM-DSSA, the first 206 scores explained 90% variance of augmented data from each variable, which is also obtained by using equation (4.3.4) and by the spectrum of the logarithm of eigenvalues. Figures 5.3–5.7 show the $\boldsymbol{T}^2$ and *SPE* charts for MM-DSSA, MM-SSA, PCA, MM-KDSSA and MM-NLSSA respectively and the results are summarised in Table 5.1.

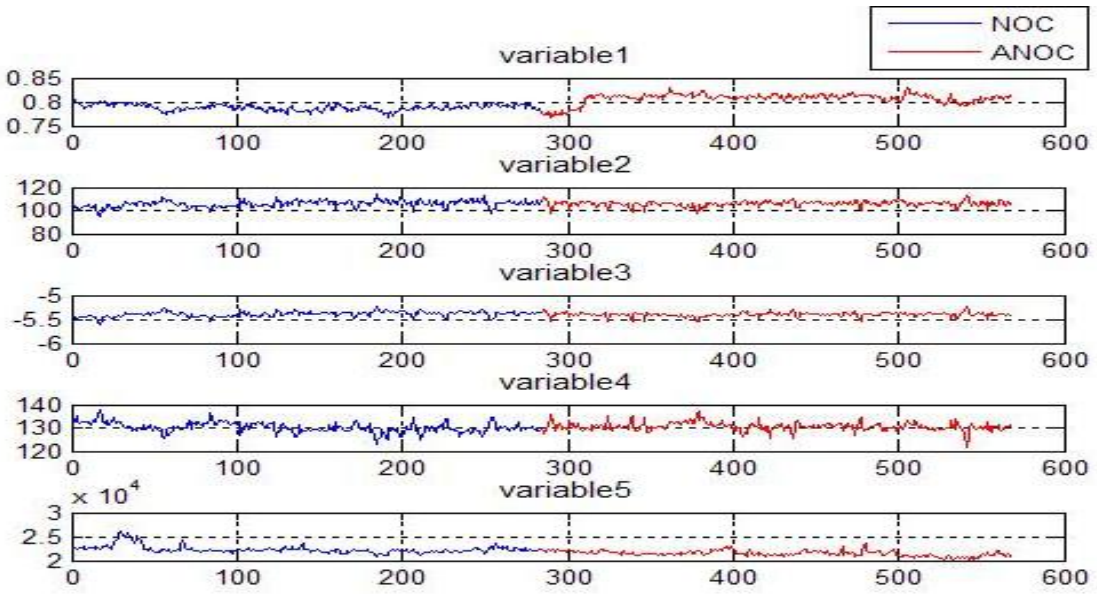


Figure 5.2: Variables from normal operating conditions (NOC) and abnormal operating conditions (ANOC) in the base metal flotation plant. In each plot the horizontal axis represents the sample size of each variable and the vertical axis represents the corresponding values of each variable.

Table 5.1: Reliability percentage of MM-KDSSA, MM-DSSA, MM-NLSSA, MS-SSA and PCA in base metal flotation plant

Method	Monitoring statistics	Case0	Case1
PCA	T^2	7.04%	4.9%
	Q	6%	94.01%
MS-SSA	T^2	2.5%	67.96%
	Q	2.1%	98.94%
MM-NLSSA	T^2	4.92%	89.43%
	Q	3.52%	57.39%
MM-DSSA	T^2	7.75%	96.65%
	Q	8.1%	85.21%
MM-KDSSA	T2	6.33%	100%
	Q	4.23%	100%

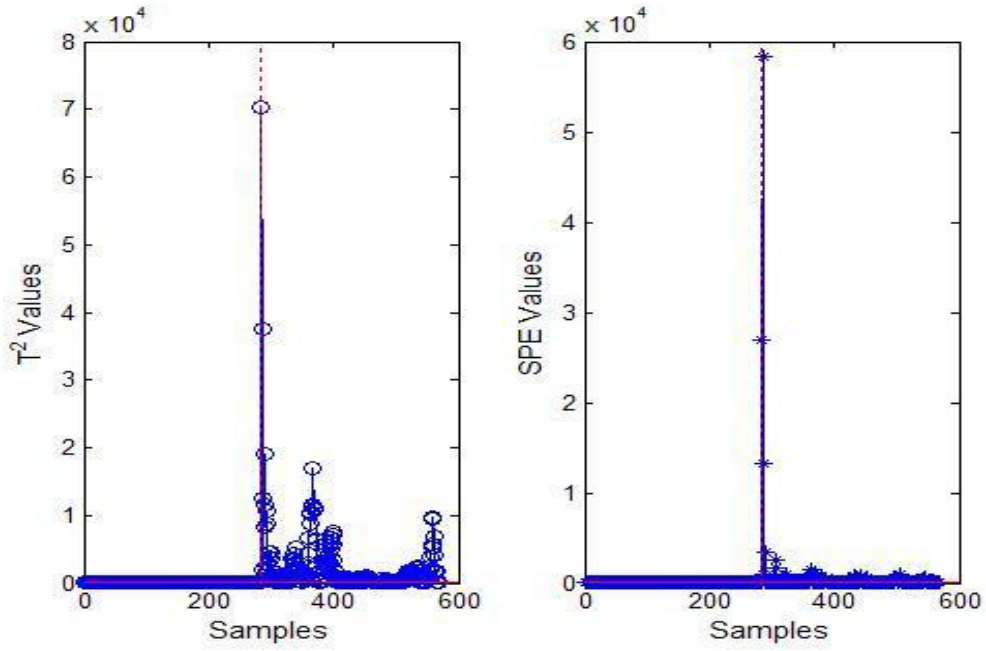


Figure 5.3: The T^2 and SPE chart for MM-DSSA in a base metal flotation plant with 95% confidence limit for both normal data (first 284 samples) and the faulty data (last 284 samples). MM-DSSA detected the faut in T^2 chart in mode 123 and SPE chart in mode 128 with highest reliability percentage.

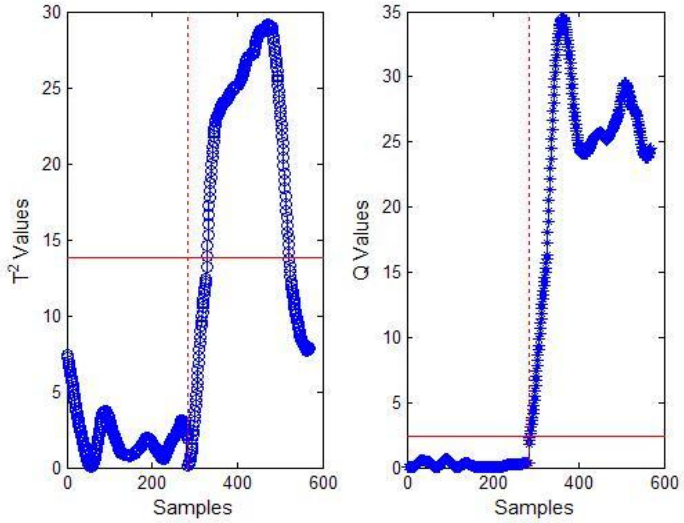


Figure 5.4: The T^2 and SPE chart for MS-SSA in base metal flotation plant with 95% confidence limit for both normal and faulty data. MS-SSA detected the fault in SPE chart in mode 1 with a 98.94% reliabilty whereas almost half of the faulty samples in T^2 chart of mode 1 is undected in this study.

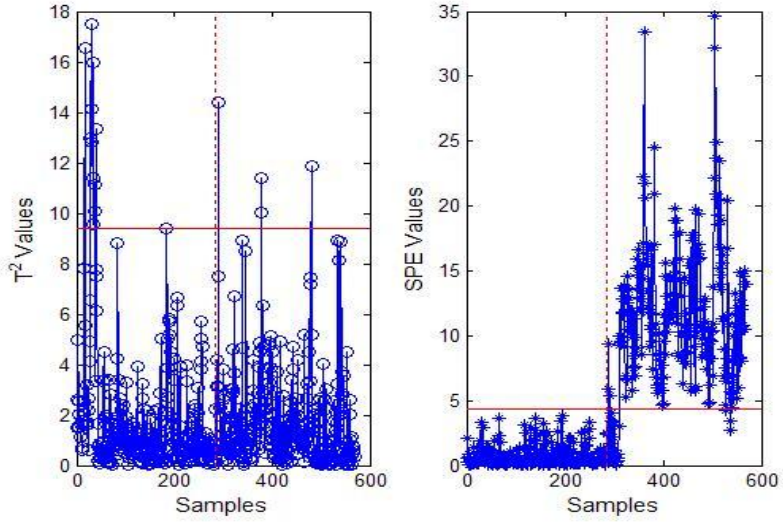


Figure 5.5: The T^2 and SPE chart for PCA in base metal flotation plant with a 95% confidence limit for both normal and faulty data. In PCA, 95% of faulty samples remain undetected in T^2 chart while all the faulty samples exceeded the control limits in the SPE chart.

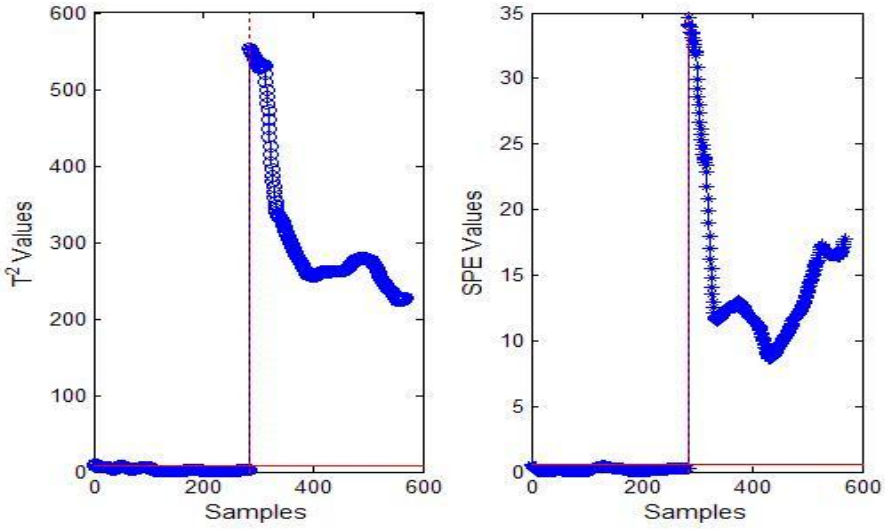


Figure 5.6: The T^2 and SPE chart for MM-KDSSA in a base metal flotation plant with a 95% confidence limit for both normal and faulty data. All the faulty samples in mode 1 of the T^2 and SPE chart exceeded the control limits in this study.

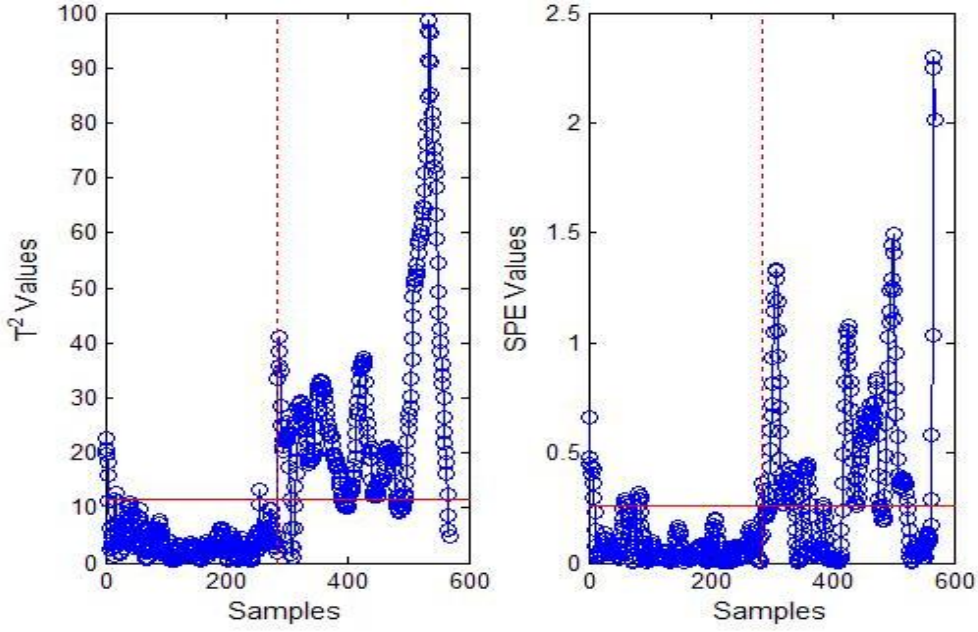


Figure 5.7: The T^2 and SPE chart for MM-NLSSA in a base metal flotation plant with a 95% confidence limit for both normal and faulty data. Most of the faulty samples in T^2 chart in mode 1 violated the control limit while almost half of the samples remains undetect in mode 2 of the SPE chart.

All the methods adopted a 95% control limit for the monitoring charts and have a reliability percentage close to 5% in Case 0 (normal operating condition) in both $\boldsymbol{T}^2$ and SPE charts, which indicated the good approximation for control limits of the $\boldsymbol{T}^2$ and SPE chart. The results from Figure 5.6 and Table 5.1 show that MM-KDSSA performed better in detection of faults with the highest reliability in both statistics in the first mode. Whereas for MM-DSSA better detection was in mode 123 and 128 for MM-NLSSA it was in the first and second mode and in MS-SSA it was obtained in the first mode for $\boldsymbol{T}^2$and SPE statistics. However, it is interesting to notice that in MM-KDSSA and MM-DSSA based monitoring both $\boldsymbol{T}^2$ and SPE statistics gave the highest reliability while in PCA and MS-SSA the highest reliability was obtained only in SPE statistics, whereas in MM-NLSSA, the reliability obtained by $\boldsymbol{T}^2$ statistics is higher than SPE statistics. Hence it can be concluded that the systematic variation in the PC subspace that changes the correlation between the variables and the respective change in the operating region is captured well in MM-KDSSA and MM-DSSA compared to PCA, MM-SSA and MM-NLSSA based monitoring. MM-NLSAA requires a large amount of data for creating a reference model as done in the previous studies. This degrades the monitoring performance of MM-NLSSA in this case. Nevertheless, the performance of MM-NLSSA was comparable to that of MS-SSA in this study. In contrast, the $\boldsymbol{T}^2$ monitoring chart of PCA in Figure 5.5 barely captures the fault in this study, while the time-series patterns of the $\boldsymbol{T}^2$ and SPE statistics of multimodal methods based on SSA are sufficiently different between normal and faulty samples to indicate the change in the process (the process is in abnormal conditions).This shows the advantage of using SSA in process monitoring by decomposing the data into multiple modes with data adaptive basis function.

In summary, the results in Table 5.1 show that the reconstructed components of MM-KDSSA and MM-DSSA capture the nonlinear relationship in the process variables using dissimilarity measures better than MM-NLSSA and MM-SSA by covariance structure of the data in both statistics. Finally, the fault detection in MM-KDSSA is obtained with the extraction of smaller number of modes as found in the previous studies, which is less time consuming and more effective than other multimodal methods. In MM-DSSA and MS-SSA a larger number of modes is used to extract the components into different levels and to monitor for fault detection, which is computationally expensive and time consuming. Also MM-NLSSA suffers from the issue of sample size as indicated earlier

and optimisation issue for the selection of neurons in the neural network for the better extraction of nonlinear components.

5.2 CASE STUDY 2: TENNESSEE EASTMAN PROCESS

The Tennessee Eastman Process (TEP) is a simulation of an actual chemical process developed as a realistic industrial process for plant-wide process control problems including process monitoring and fault diagnosis (Downs & Vogel, 1993; Kano *et al*., 2002). The process consists of five major units: an exothermic two-phase reactor, a product condenser, a recycle compressor, a vapor-liquid separator, and a product stripper. It also involves eight components A, B, C, D, E, F, G, and H. G and H components are liquid products produced from the gaseous reactants A, C, D, and E. The component F is an undesired byproduct. The process flow diagram is shown in the Figure 5.8.

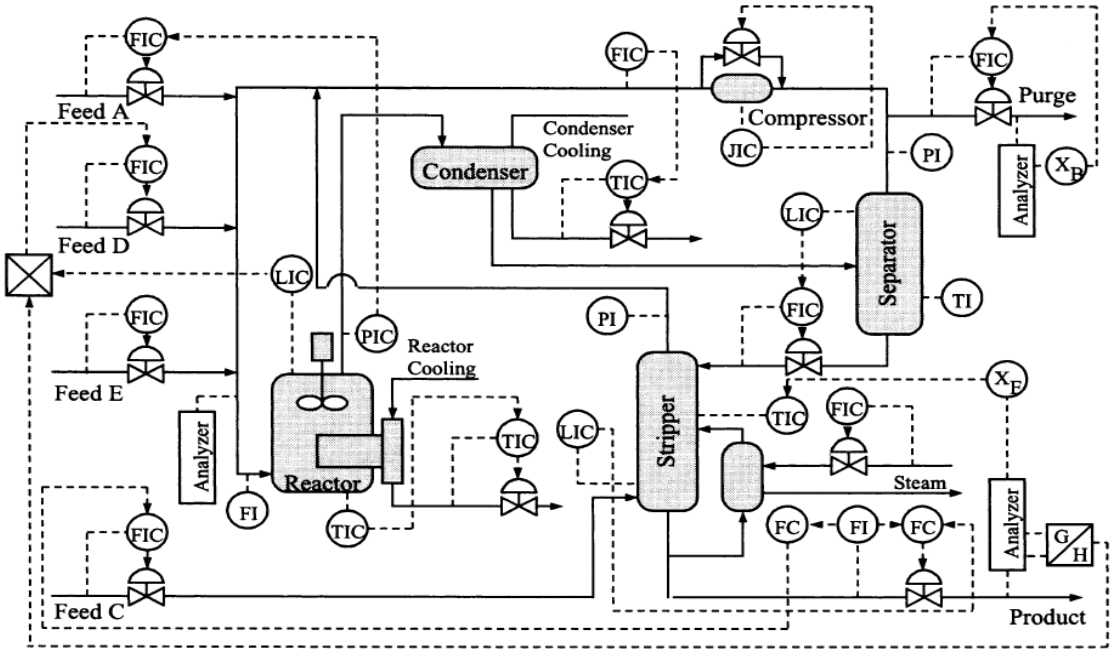


Figure 5.8: Process flow sheet for TEP (Downs & Vogel, 1993).

The reactor reactions are represented by the following equations:

$$A(g) + C(g) + D(g) \rightarrow G(liq), product1 \qquad \textbf{(5.2.1)}$$

$$A(g) + C(g) + E(g) \rightarrow H(liq), product2 \qquad \textbf{(5.2.2)}$$

$$A(g) + E(g) \rightarrow F(liq), byproduct \quad \textbf{(5.2.3)}$$

$$3D(g) \rightarrow 2F(liq), byproduct \quad \textbf{(5.2.4)}$$

The reactions in equation (5.2.1- 5.2.4) are exothermic, irreversible and approximately first-order with respect to the reactant concentrations. The Arrhenius function of temperature is used to represent the reaction rates. The reaction for G product formation in equation (5.2.1) has higher activation energy than the reaction for H in equation (5.2.2) and therefore, results in higher sensitivity to temperature.

The liquid products are formed by the reactant gases, which are catalysed by a non-volatile catalyst dissolved in the liquid. The cooling water circulated in the reactor reduced the heat of the reaction. The catalyst is retained while the gaseous products are separated from the reactor. The product gas is then cooled by a condenser after which it is fed to a vapor-liquid separator. By the use of a compressor, the non-condensed vapor from the separator is recycled to the reactor. A portion of the recycled vapor is purged to keep the inert product and byproduct from accumulating in the process using the vapor-liquid separator. The reactants that remain in the condensed stream from the separator are then removed in the stripper.

The process provides 12 manipulated or control variables and 41 measured or observed variables. Of the 41 measured variables, 22 are continuous and 19 are composition measurements. Sixteen variables from the 53 process variables are used, as selected by Chen & McAvoy (1998), for monitoring purposes and are listed in Table 5.2.

Table 5.2: Process variables used for monitoring.

Variable	Description
1	A feed
2	D feed
3	E feed
4	A and C feed
5	Recycle flow
6	Reactor feed rate
7	Reactor temperature
8	Purge rate
9	Product separator temperature
10	Product separator pressure
11	Product separator underflow
12	Stripper pressure
13	Stripper temperature
14	Stripper steam flow
15	Reactor cooling water outlet temperature
16	Separator cooling water outlet temperature

The TEP process contains 21 fault cases, which are listed in Table 5.3, of which 16 are known and 5 are unknown. Each of the first 7 faults are associated with a step change in a process variable while faults 8 to 12 represent the increased variability in some of the process variables. Fault 13 is associated with a slow drift in the reaction kinetics. Faults 14, 15, and 21 are associated wih actuator faults such as sticking valves.

Table 5.3: Process faults

Case	Disturbance	Type
1	A/C feed ratio, B composition constant	Step
2	B composition, A/C ratio constant	Step
3	D feed temperature	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	A feed loss	Step
7	C header pressure loss - reduced availability	Step
8	A, B, C feed composition	Random variation
9	D feed temperature	Random variation
10	C feed temperature	Random variation
11	Reactor cooling water inlet temperature	Random variation
12	Condenser cooling water inlet temperature	Random variation
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16-20	Unknown	Unknown

The data representing normal operating conditions and the 21 fault cases contained 500 samples in this study. The monitoring performance of MM-KDSSA, MM-DSSA and MM-NLSSA methods for all 21 fault conditions were evaluated to compare the proficiency of multimodal nonlinear methods in the respective fault condition and the monitoring results are also compared to MS-SSA and PCA. As in the previous case studies, 95% control limits are set for both $\boldsymbol{T}^2$and SPE statistics in all methods. The reliability percentage was calculated from samples that violate the 95% confidence limit in respective modes of each monitoring method. In each mode the ratio of number of samples violating the control limits in both $\boldsymbol{T}^2$ and SPE values were calculated, and the maximum ratio was taken as the reliability percent of the respective fault condition (as summarised in Table 5.5). The reliability in Case 0 is considerably higher or close to 5% in all the methods, which indicates that the control limits are determined successfully to detect the faults in all the models. A detailed discussion on the monitoring performance of nonlinear SSA in some of the difficult fault conditions are evaluated with performance of MS-SSA and PCA methods based on the reliability percentage in Table 5.5.

The monitoring performances of MM-KDSSA, for all 21 fault conditions were evaluated with a window size of M=38. The embedding dimension M was selected based on the first maximal decorrelation point of monitored variables by equation (3.2.1) and was

used for computing the trajectory matrices of 16 variables. The selection of the eigenvalues and eigenvectors to represent the reconstructed components in multiple modes were determined based on equation (4.3.4). The logarithm of the singular values of the doublecentered kernel matrices obtained from the dissimilarity matrix of lagged trajectory matrix of the normal data accounted for at most 99% variation of the augmented data from each variable. A drop in the singular value occurring at around the third singular value indicates the start of the noise component and hence, the first three modes from 463 modes were used to build the PCA model for each fault condition. The PCs retained in the PCA model of reconstructed signal in each mode captured at least 90% variance in the data. The reliability percentage for $\boldsymbol{T}^2$ and SPE obtained by the MS-KDSSA model for 21 fault conditions are summarised in Table 5.5 based on 95% control limits for monitoring statistics.

In MM-NLSSA, a window size of M=38 is used for SSA decomposition of 16 variables. The number of PCs retained in SSA decomposition for each variable as the input to the NLPCA.cir network explained at least 90% of variance in the data. The number of hidden neurons for NLPCA.cir network for each variable was determined based on the minimum mean square error *m.s.e* between the neural network output $\boldsymbol{T}'$ and the original PC $\boldsymbol{T}$. The number of PCs retained in SSA decomposition and the numbers of chosen neurons for the NLPCA.cir network for each variable are listed in Table 5.4. The PCs retained in the PCA model of reconstructed signal in each mode, capture at least 96% variance in the data. The maximum reliability percentage for $\boldsymbol{T}^2$ and SPE obtained from the first two modes for all 21 fault conditions are shown in Table 5.5 as indicated before.

Table 5.4: Number of PCs retained in SSA decomposition and the number of hidden neurons used in NLPCA.cir for 16 variables

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Number Of PCs	26	3	3	3	3	3	3	19	3	3	28	3	3	3	3	3
Number Of Neurons	22	5	6	5	4	3	5	5	5	4	10	4	7	7	4	5

An embedding window size of M=38 was determined based on the first maximal decorrelation point of monitored variables, leading to a 38-level multimodal representation in the case of MS-SSA. The PCs retained in the PCA model for each mode captured at least 96% variance in the data. The reliability percentage based on MS-SSA is calculated from the samples, which violated the 95% confidence limits in the respective modes. In each mode the ratio of the number of samples violating the control limits in both $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics values are calculated. The maximum detection rate is considered as the per cent reliability of the associated fault condition, which is listed in Table 5.5.

The performance of the MM-DSSA method for all 21 fault conditions was evaluated with a window size of M =38 for the computation of a lagged trajectory matrix. The first 145 modes from 463 modes were used to build the PCA model for each fault condition based on the logarithmic distribution of eigenvalues and equation (4.3.4). The PCs retained in the PCA model of reconstructed signal in each mode, capture at least 90% variance in the data. The reliability percentage for $\boldsymbol{T}^2$ and SPE obtained by MM-DSSA for 21 fault conditions are listed in Table 5.5. The control limits for both statistics were set at 95% confidence level in all cases as indicated previously.

Using cPCA, 13 PCs explaining 96% of the variation in the data for the normal case were retained in the PCA model. The estimated reliability of $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics in each case using cPCA are also listed in Table 5.5, as mentioned earlier.

Table 5.5: Reliability (%) of MS-SSA, PCA, MM-NLSSA, MM-DSSA and MM-KDSSA in TE Process

	Process monitoring methods									
Cases	**PCA**		**MS-SSA**		**MM-DSSA**		**MM-KDSSA**		**MM-NLSSA**	
	T^2	Q	T^2	Q	T^2	Q	T^2	Q	T^2	Q
Case 0	2	8	1	0	5	4	7.2	4.2	1.4	5.8
Case 1	100	100	100	100	100	100	100	100	100	98
Case 2	100	93	100	100	100	100	100	100	100	98
Case 3	60	16	48	51	100	99	100	100	100	86
Case 4	29	9	31	79	100	99	100	100	100	83
Case 5	100	81	100	100	100	100	100	100	100	90
Case 6	100	100	100	100	100	100	100	100	100	97
Case 7	100	87	100	100	100	100	100	100	100	89
Case 8	98	90	100	100	100	100	100	100	100	93
Case 9	11	10	31	46	100	99	100	100	100	92
Case 10	84	86	93	95	100	99	100	100	100	92
Case 11	67	14	93	86	100	99	100	100	100	96
Case 12	100	89	100	100	100	100	100	100	100	95
Case 13	83	88	100	100	100	99	100	100	100	90
Case 14	100	93	98	65	100	100	78.2	100	100	90
Case 15	8	7	37	49	100	100	100	100	100	93
Case 16	28	12	55	87	100	100	100	100	100	79
Case 17	99	95	100	100	100	100	100	100	100	86
Case 18	52	48	68	87	100	99	100	100	100	93
Case 19	40	14	72	91	100	100	100	100	100	99
Case 20	45	32	54	81	100	99	100	100	100	93
Case 21	9	10	20	100	100	99	100	100	100	92

The reliability of MM-KDSSA is higher (100%) than that of PCA, MS-SSA, MM-DSSA, and MM-NLSSA for all fault conditions in both statistics. More specifically, the fault case 1-2, case 5-8, case 10, case 12-13 and case 17 are easily detected by all methods and some of them are discussed briefly below. Almost all the methods have above 80% reliability percentage in both $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics. For case 14, all the methods offer high reliability (90%-100%) in $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics except that MS-SSA and MM-KDSSA have 65% and 78.2% for $\boldsymbol{Q}$ and $\boldsymbol{T}^2$ statistics respectively. For fault 1 (Case 1) there is a step change in the ratio of A/C feeds causing variations in flow rates and compositions of stream 6 and 4. Many variables in the reactor are affected by these variations and hence, the fault is detected easily by all methods. In fault 2 (Case 2), the step change of gaseous inert B in stream 4 caused an increase in the stripper pressure which leads to

variations in compositions in the reactor feed, the purge stream and the product stream. Many variables, therefore, behave differently from the normal case (Case 0) and hence, the fault is detected easily by all the methods.

In Fault 5 the step change in the condenser cooling water inlet temperature caused an increase in the flow rate of the outlet stream from the condenser to the vapor-liquid separator. This resultsed in an increase in the gas-liquid temperature and the separator cooling water outlet temperature. In this case, SSA based methods gave 100% reliability in both $\boldsymbol{T^2}$ and $\boldsymbol{Q}$ statistics while PCA gave 100% reliability in $\boldsymbol{T}^2$ and 81% reliability in $\boldsymbol{Q}$ statistics. This is because the combination of SSA based analysis and PCA extracts richer information from correlated variables and hence results in a better detection rate.

Fault 6 is associated with component A feed loss in stream 1, which causes an increase in the reactor pressure. The effect of this fault easily propagates to most of the variables and can be detected easily. The step change in the C header pressure loss-reduced availability in fault 7 and random variation in the temperature of stream 4 (C feed temperature) in fault 10 cause variations in the process variables. This resulted in getting good detection reliability with most of the monitoring methods. Similarly, variations in the relationships of the process variables in fault 8, 12, 13 and 17 also gave high detection rate with most of the methods in this study.

In fault 14, the reactor cooling water valve was stuck. Most of the methods could detect this fault effectively with high reliability percentage. However, the $\boldsymbol{Q}$ statistics of MS-SSA and $\boldsymbol{T}^2$ statistics of MM-KDSSA performed poorly when compared to the other methods in those statistics.

For fault cases 4, 11, 16, 18, 19, 20 and 21, SSA based multimodal methods provided improved performance monitoring over the standard PCA. In particular, SSA based nonlinear methods provide superior results in the detection of these faults with a high reliability percentage compared to MS-SSA. The root causes for faults 16 to 20 are unknown and are not easy to detect. Therefore, the performance of conventional MSPC methods on these faults will have obvious differences. A detailed discussion on the fault detection performance using $\boldsymbol{T}^2$and $\boldsymbol{Q}$ charts are given below for some of the selected faults in the above cases.

The Case 4 occurs due to a sudden temperature increase in the reactor. This causes a step change reduction in the reactor cooling water inlet temperature. Hence, the temperature increases quickly but small variations in the measurements compared to the normal operating conditions may occur as a result of a change in the temperature of the reactor. MM-KDSSA, MM-DSSA and MM-NLSSA detected these small events better than the case with PCA and MS-SSA. MM-KDSSA and MM-NLSSA detected the faults in the first mode while MM-DSSA detected the fault at the mode 116 as shown in Figures 5.9-5.13. Similarly, fault condition 11 was associated with random variation in the reactor cooling water inlet temperature. Hence, large oscillations in the reactor cooling water flow rate induce fluctuations in the reactor temperature. The step change in D feed temperature causes a mean shift in flow and outlet temperature of the reactor cooling water via a cascade control system for reactor temperature (Kano *et al*., 2002). This small mean shift in the measurement caused a change in the relationship between the process variables as defined under normal operating conditions. In contrast to PCA and MS-SSA, both $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics of MM-DSSA, MM-KDSSA and MM-NLSSA in Figures 5.14-5.18 showed high performance in detection of these faults. This shows the advantage of using nonlinear methods, MM-KDSSA, MM-DSSA and MM-NLSSA in detecting mean shifts in the process compared to linear methods such as conventional PCA and MS-SSA. This is achieved by monitoring the reconstructed components in the first mode of MM-KDSSA and MM-NLSSA while MM-DSSA gave better performance in mode 42.

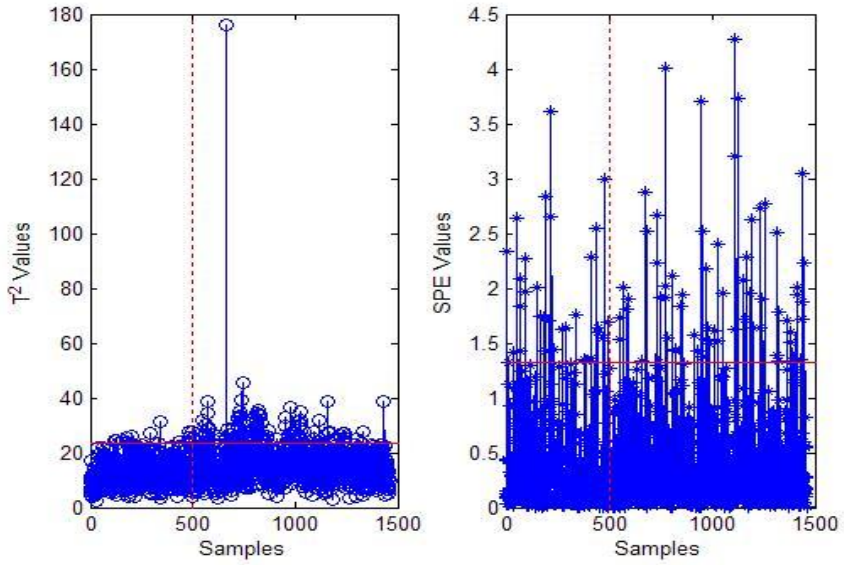


Figure 5.9: The T^2 and SPE chart for PCA in fault condition 4 for both normal data (first 500 samples) and the fault condition data (last 500 samples) with 95% confidence limit.

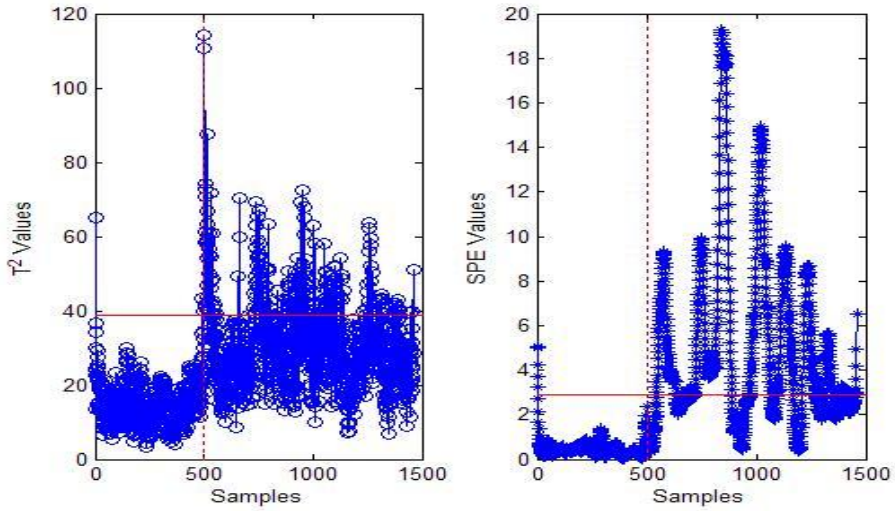


Figure 5.10: The T^2 and SPE chart for MS-SSA in fault condition 4 for both normal and abnormal data with 95% confidence limit.

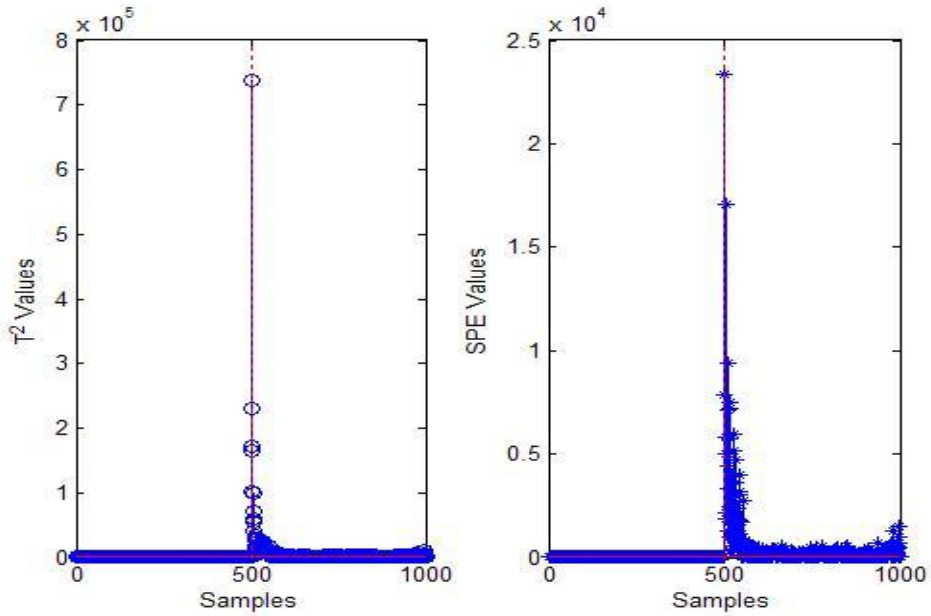


Figure 5.11: The T^2 and SPE chart for MM-DSSA in fault condition 4 for both normal and abnormal data with 95% confidence limit.

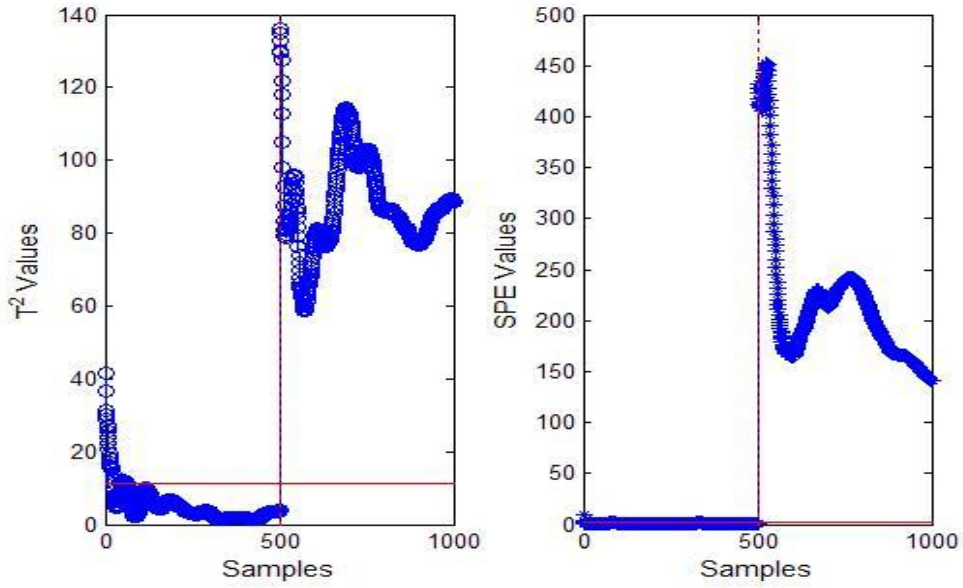


Figure 5.12: The T^2 and SPE chart for MM-KDSSA in fault condition 4 for both normal and abnormal data with 95% confidence limit.

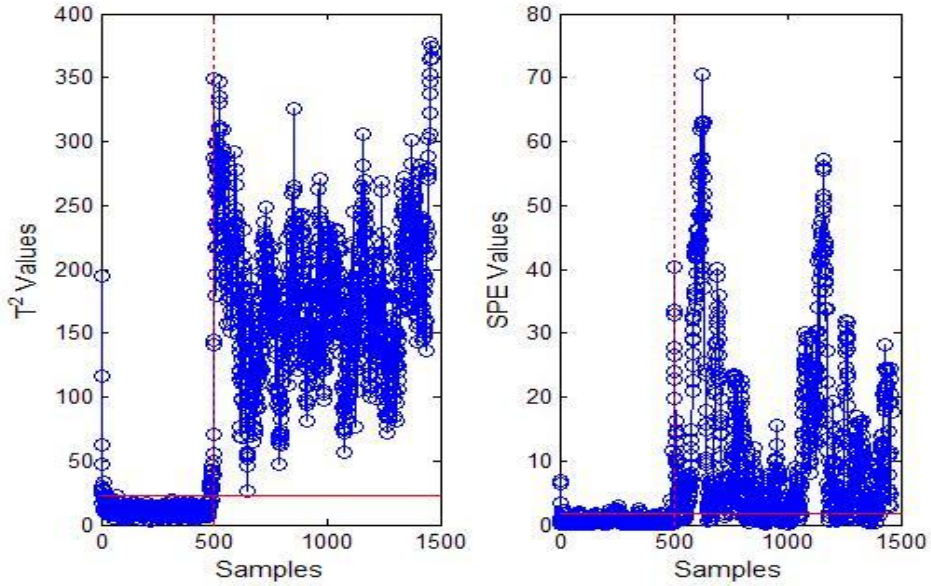


Figure 5.13: The T^2 and SPE chart for MM-NLSSA in fault condition 4 for both normal and abnormal data with 95% confidence limit.

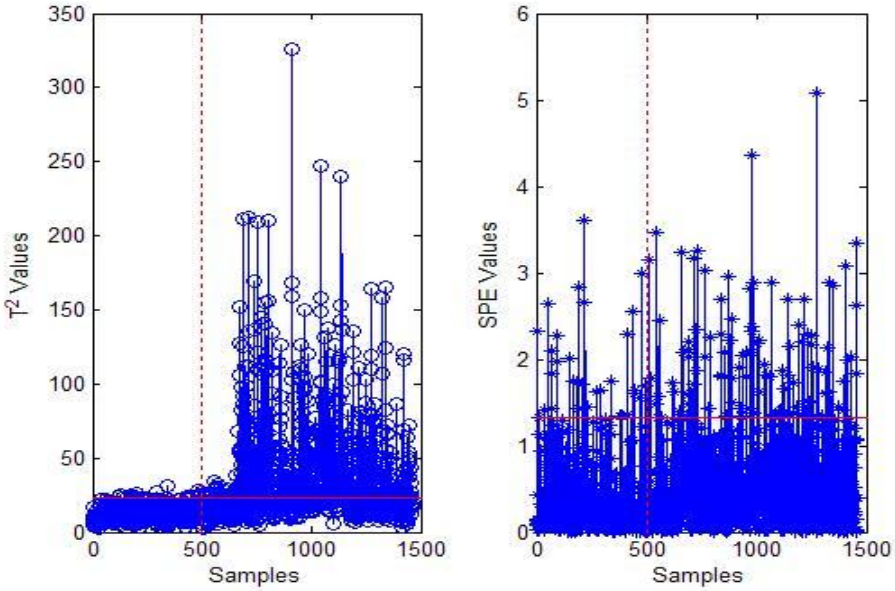


Figure 5.14: The T^2 and SPE chart for PCA in fault condition 11 for both normal and abnormal data with 95% confidence limit.

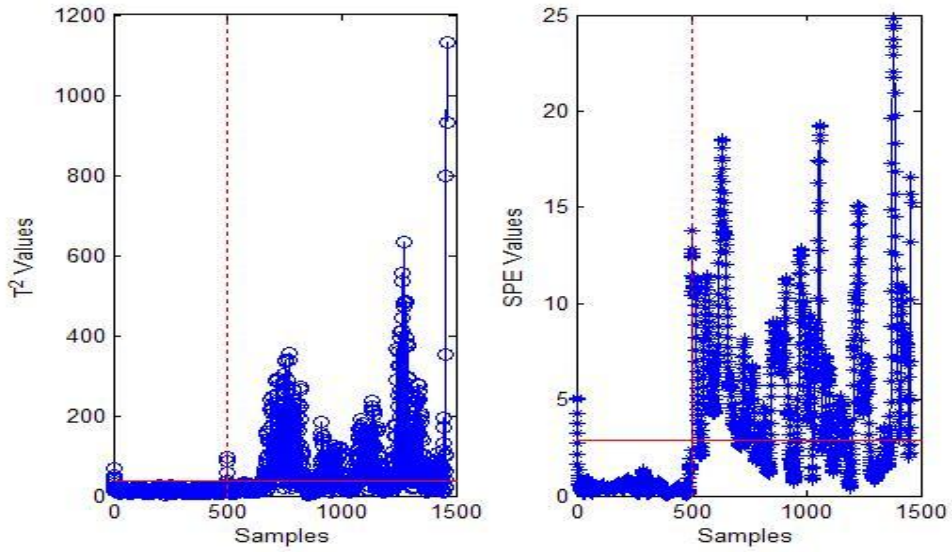


Figure 5.15: The T^2 and SPE chart for MS-SSA in fault condition 11 for normal and abnormal data with 95% confidence limit.

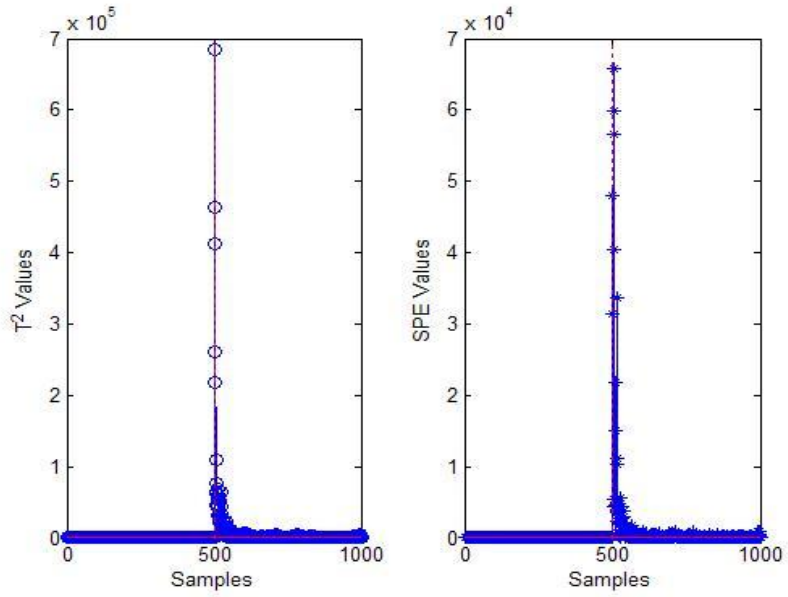


Figure 5.16: The T^2 and SPE chart for MM-DSSA in fault condition11 for both normal and abnormal data with 95% confidence limit.

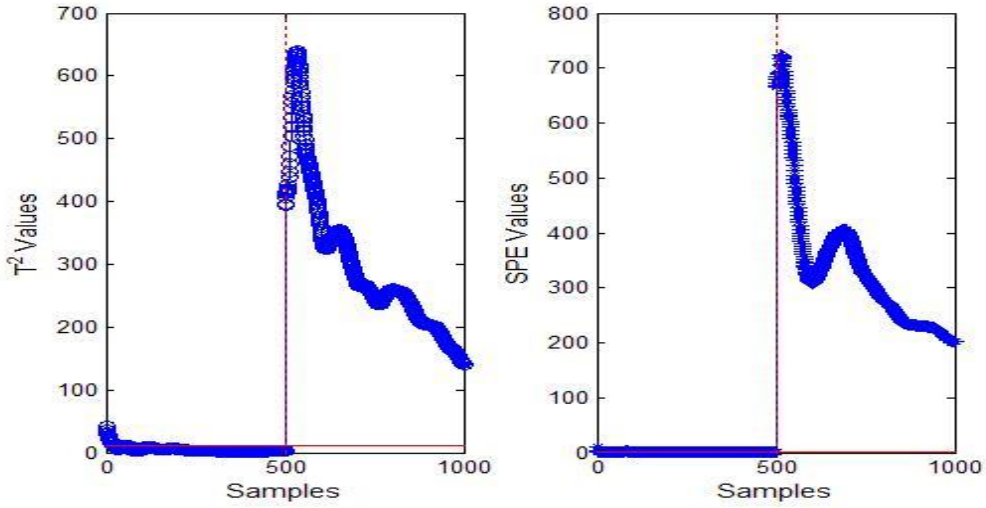


Figure 5.17: The T^2 and SPE chart for MM-KDSSA in fault condition11 for both normal and abnormal data with 95% confidence limit.

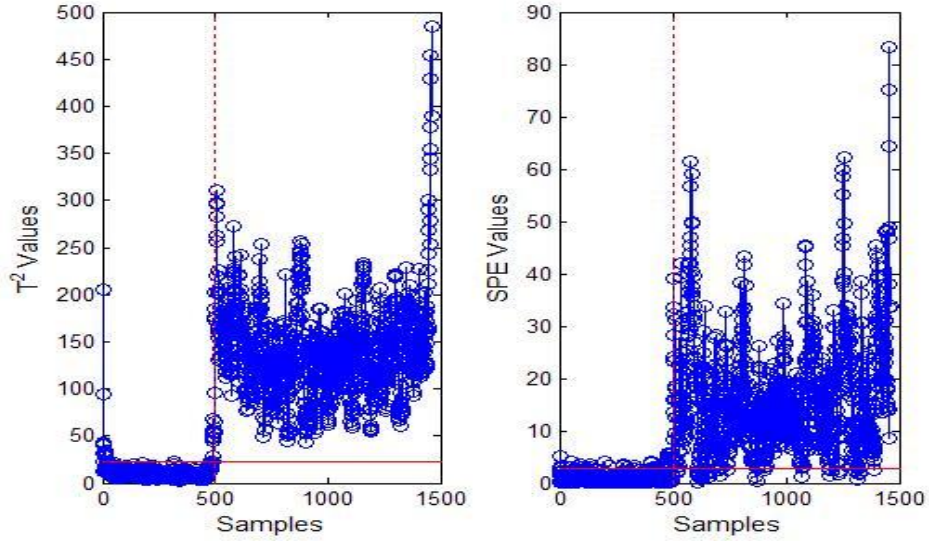


Figure 5.18: The T^2 and SPE chart for MM-NLSSA in fault condition11 for both normal and abnormal data with 95% confidence limit.

Cases 16 and 19 produced feeble and slow changes in the process that were relatively difficult to detect with classical multivariate methods. The root cause of case 16 was unknown. The detection results using $\boldsymbol{T}^2$ and $\boldsymbol{Q}$ statistics for PCA, MS-SSA, MM-NLSSA,

MM-DSSA and MM-KDSSA are summarised in Table 5.5 and are shown in Figure 5.19-5.23.

MM-KDSSA, MM-DSSA and MS-SSA showed better detection compared to PCA. This is because case16 is associated with random variation and hence, affects the residual space (Lau *et al*., 2013). The fault rich information in the residual space was filtered by data adaptive basis functions using SSA. Features are then extracted in multiple modes for better detection of change in the variables as compared to PCA in which the feature extraction was obtained in single mode. Therefore, MM-KDSSA, MM-DSSA MM-NLSSA and MS-SSA gave a high reliability percentage in $\boldsymbol{Q}$ statistics compared to PCA.

The small changes in Case 19 affected the recycle flow in TE-process and affected the residual space, which also measured the random variations in the process (Lau *et al*., 2013). The value of the $\boldsymbol{Q}$ statistics was above the threshold for most of the time for MM-KDSSA, MM-DSSA, MM-NLSSA and MS-SSA compared to PCA as shown in Figures 5.24-5.28. PCA used only the score space information, which extracted large deviations in the process variables while information in the residual space for random variations is discarded. On the other hand, MM-KDSSA, MM-DSSA, MM-NLSSA and MS-SSA had better detection of changes in the stochastic behavior of the variable by the extraction of features in multiple modes using SSA. In particular, MM-KDSSA, MM-NLSSA and MM-DSSA utilised the residual space information effectively to capture deviations in the process variables using nonlinear SSA and hence, achieved a better detection compared to MS-SSA and PCA in most of these cases in both $\boldsymbol{T}^2$and $\boldsymbol{Q}$ statistics. In all the above cases, MM-KDSSA outperformed MM-DSSA, MM-NLSSA and MS-SSA in detecting faults in relatively smaller modes (first mode) in both statistics. Moreover, the detection values for both statistics are higher in MM-KDSSA and MM-DSSA compared to MM-NLSSA and MS-SSA in all the fault conditions. This shows the advantage of using dissimilarity measure to capture the random variation in the process.

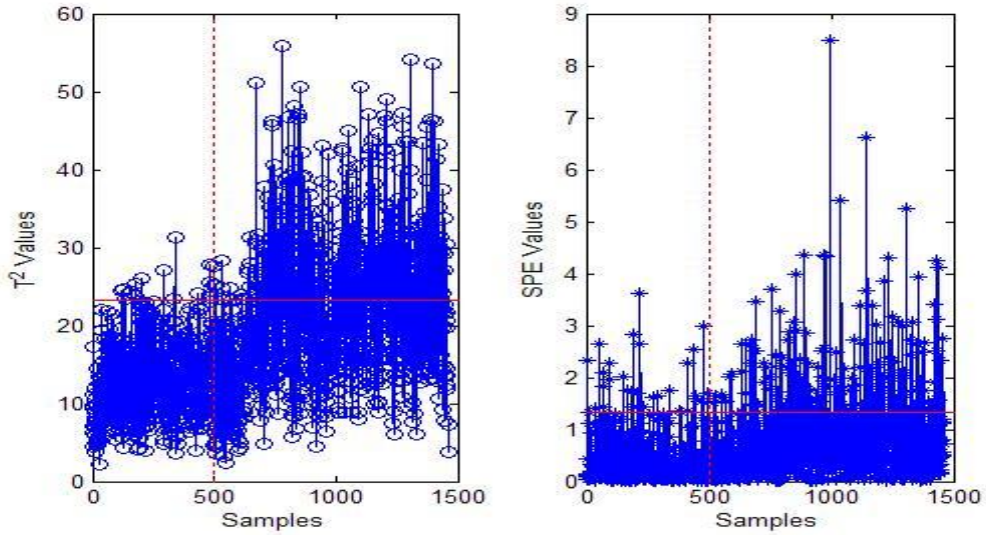


Figure 5.19: The T^2 and SPE chart for PCA in fault condition16 for both normal data (first 500 samples) and abnormal data (last 500) samples with a 95% confidence limit.

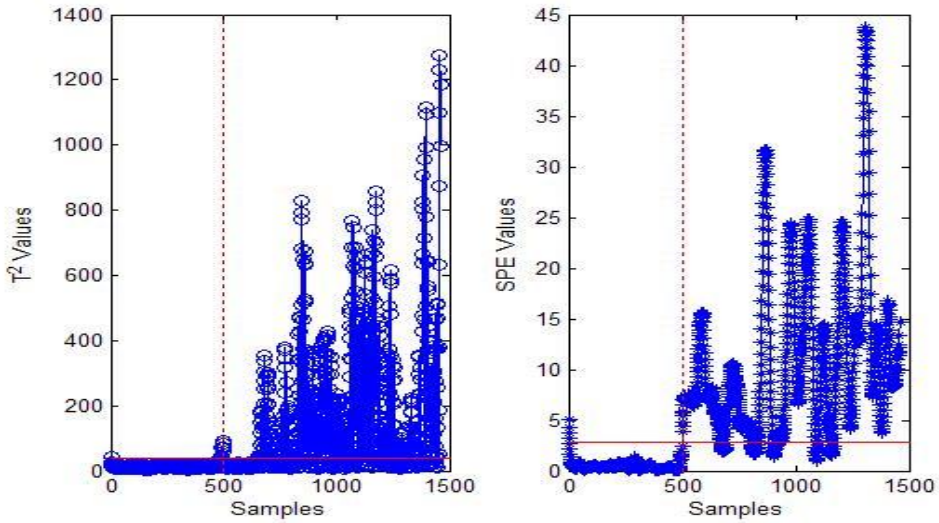


Figure 5.20: The T^2 and SPE chart for MS-SSA in fault condition16 for both normal and abnormal data with 95% confidence limit.

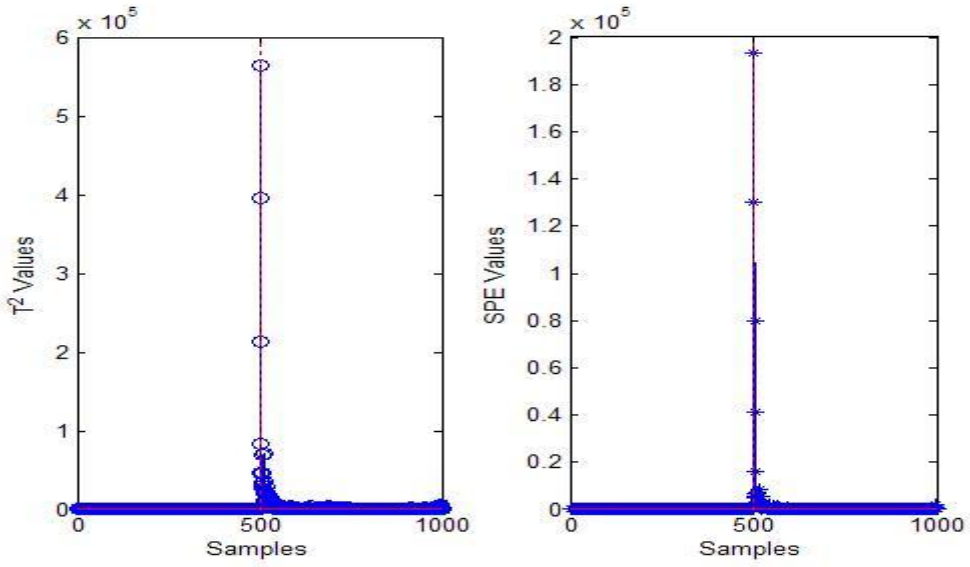


Figure 5.21: The T^2 and SPE chart for MM-DSSA in fault condition 16 for both normal and abnormal data with 95% confidence limit.

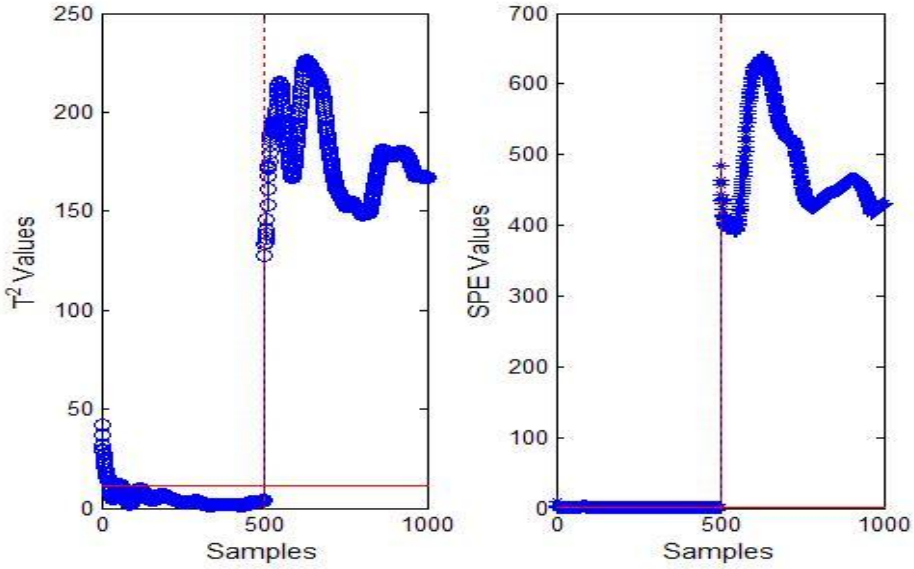


Figure 5.22: The T^2 and SPE chart for MM-KDSSA in fault condition16 for both normal and abnormal data with 95% confidence limit.

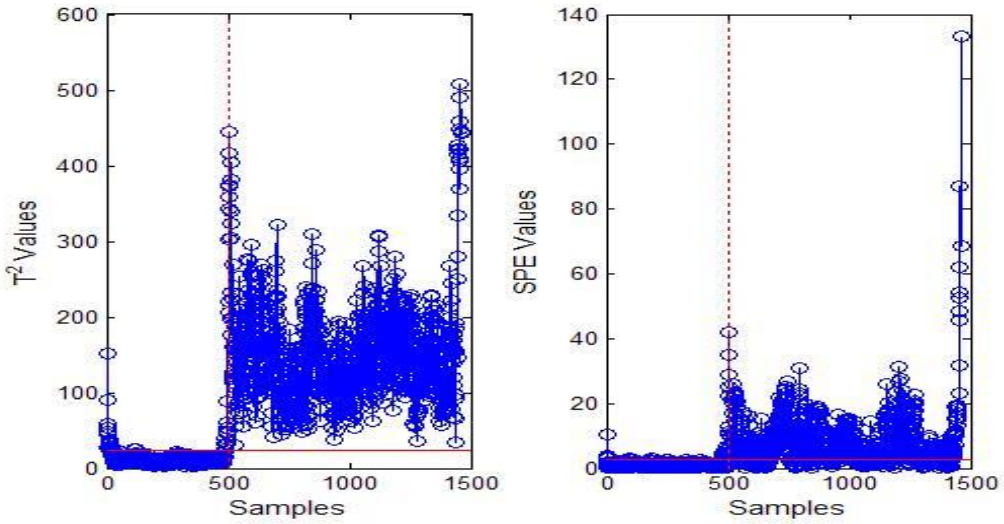


Figure 5.23: The T^2 and SPE chart for MM-NLSSA in fault condition 16 for both normal and abnormal data with 95% confidence limit.

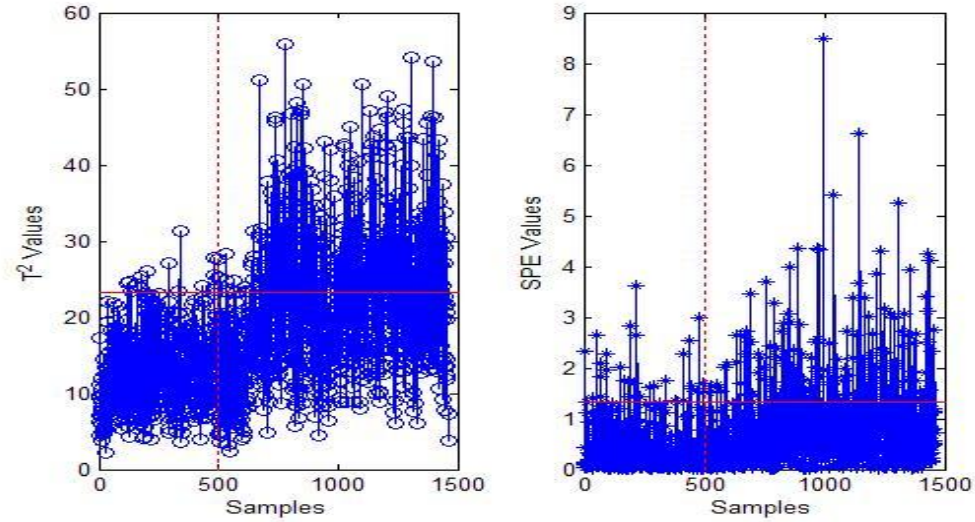


Figure 5.24: The T^2 and SPE chart for PCA in fault condition 19 for both normal and abnormal data with a 95% confidence limit.

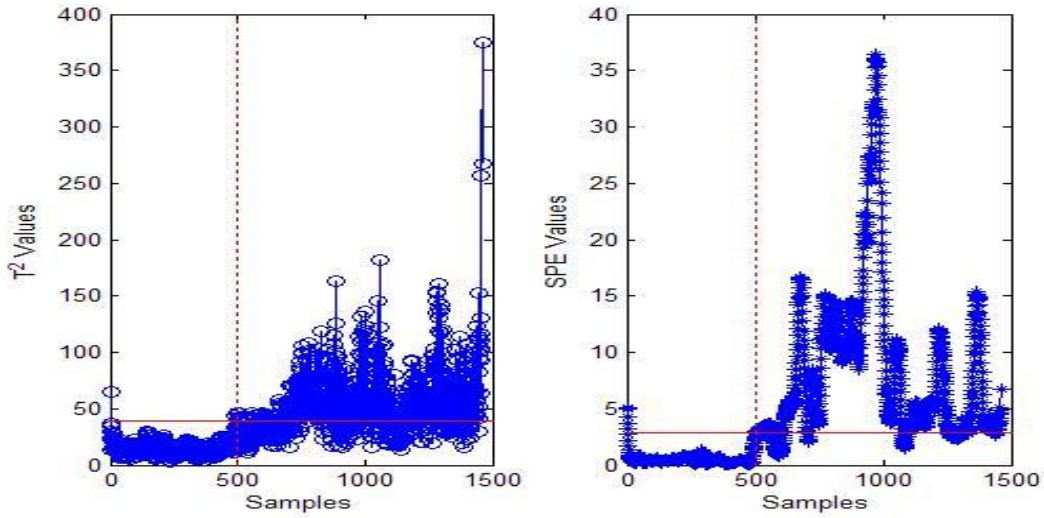


Figure 5.25: The T^2 and SPE chart for MS-SSA in fault condition 19 for both normal and abnormal data with 95% confidence limit.

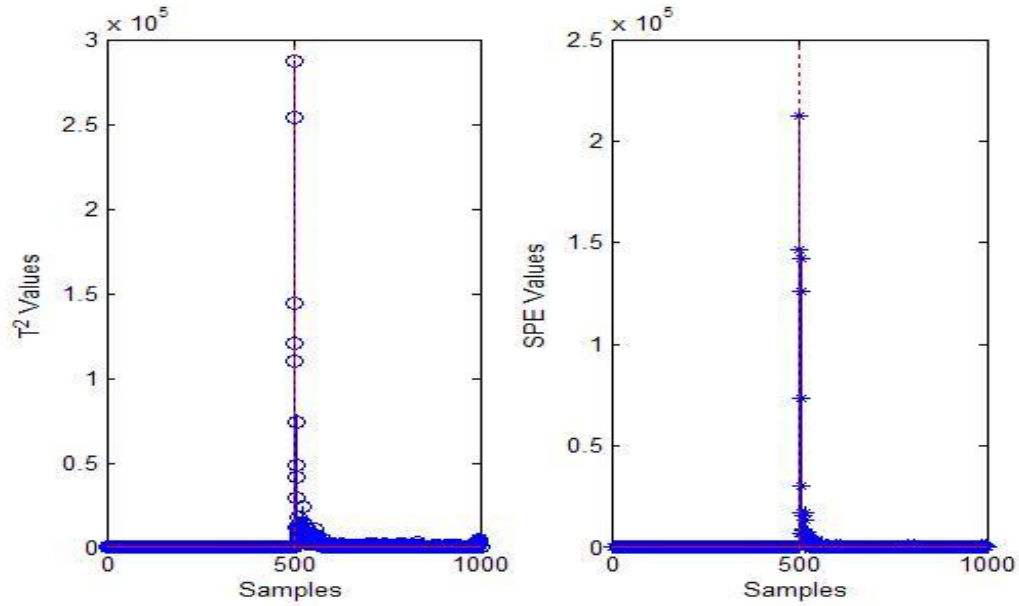


Figure 5.26: The T^2 and SPE chart for MM-DSSA in fault condition 19 for both normal and abnormal data with 95% confidence limit.

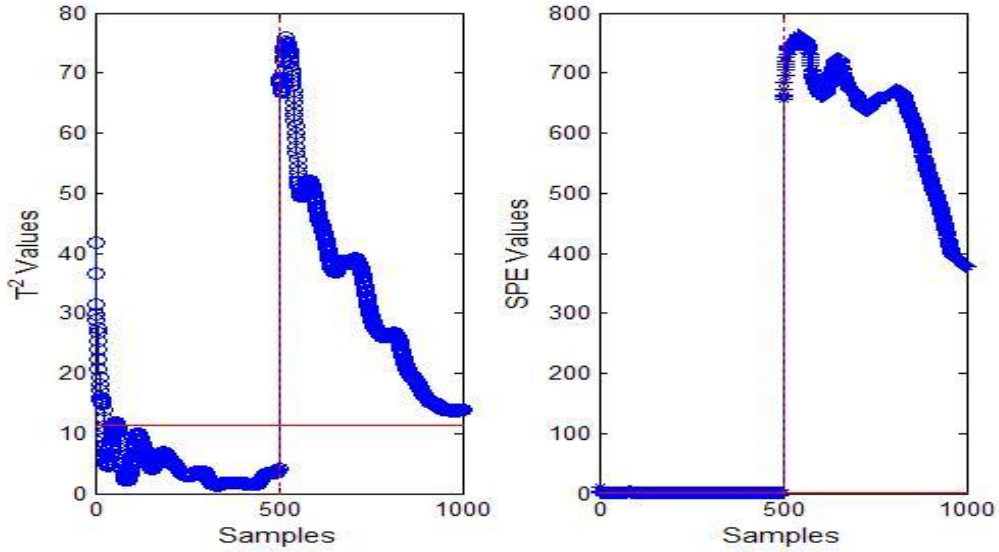


Figure 5.27: The T^2 and SPE chart for MM-KDSSA in fault condition 19 for both normal and abnormal data with 95% confidence limit.

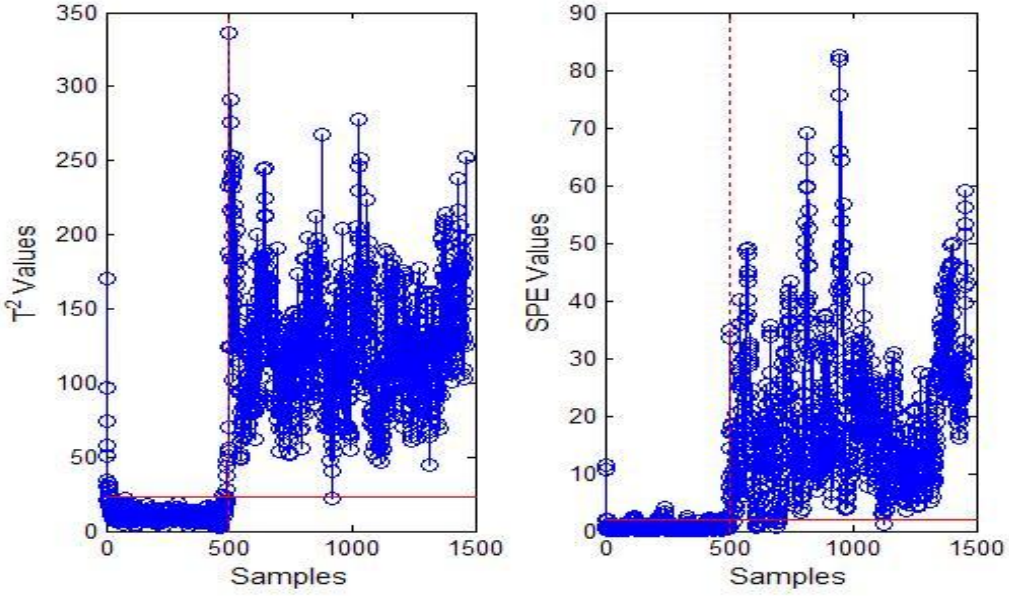


Figure 5.28: The T^2 and SPE chart for MM-NLSSA in fault condition 19 for both normal and abnormal data with 95% confidence limit.

The superiority of nonlinear, SSA-based methods for the detection of faults, which are more difficult to detect (faults 3, 9 and 15) over the other MSPC fault detection methods including MS-SSA and PCA, previously reported in the literature, is demonstrated using dissimilarity plots of process variables and plots of reconstructed components in the normal and abnormal cases of those faults. The previous works on the analysis of faults 3, 9 and 15 have found that these faults proved to be difficult for data-driven detection methods because these faults have very little effect on the corresponding process measurements. That is, these faults are defined as *unobservable* from the process data since there is no observable change in the mean or the variance that can be detected (Chiang *et al*., 2001; & Ma, 2011). In other words, variables in these faults were unobservable in the abnormal condition of TE-process as shown in Figures 5.29–5.34.

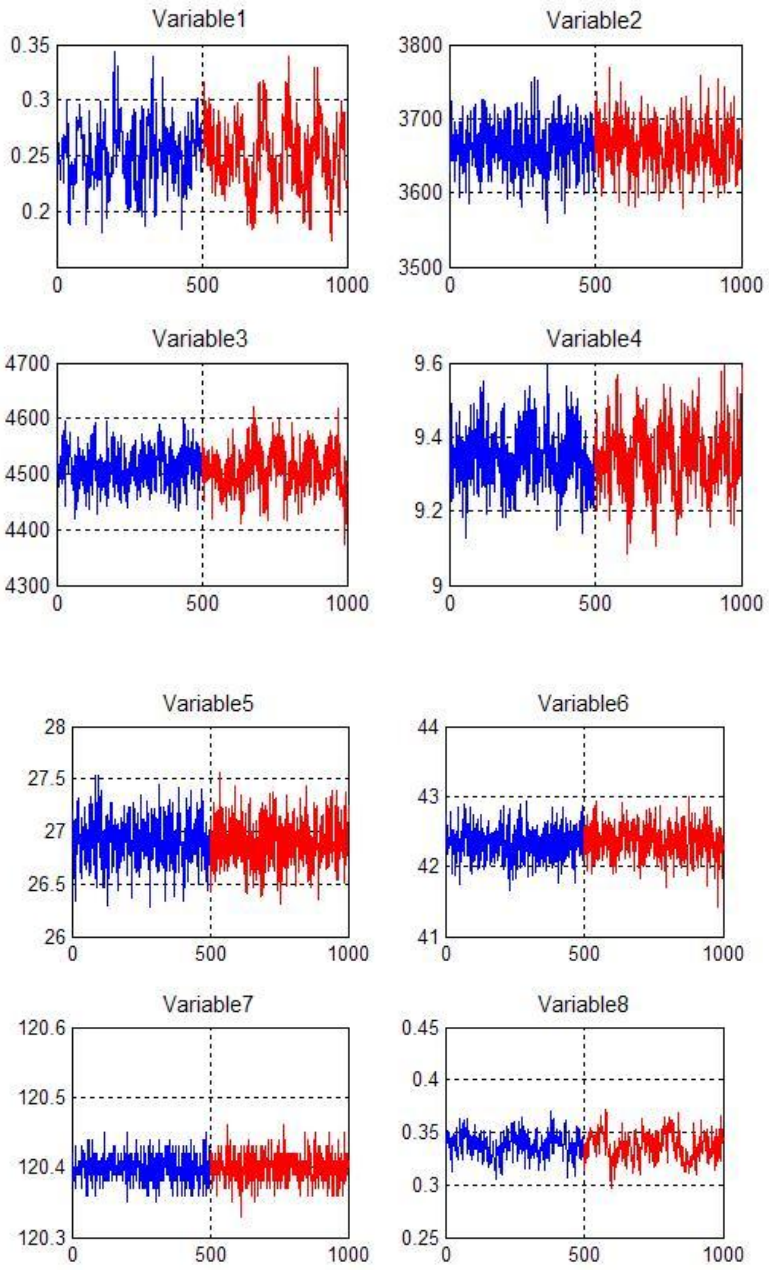


Figure 5.29: Variables (1-8) in Case 3 for normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in the TE-Process.In each plot the horizontal axis represents the sample size and the vertical axis represents the corresponding values of each variable.

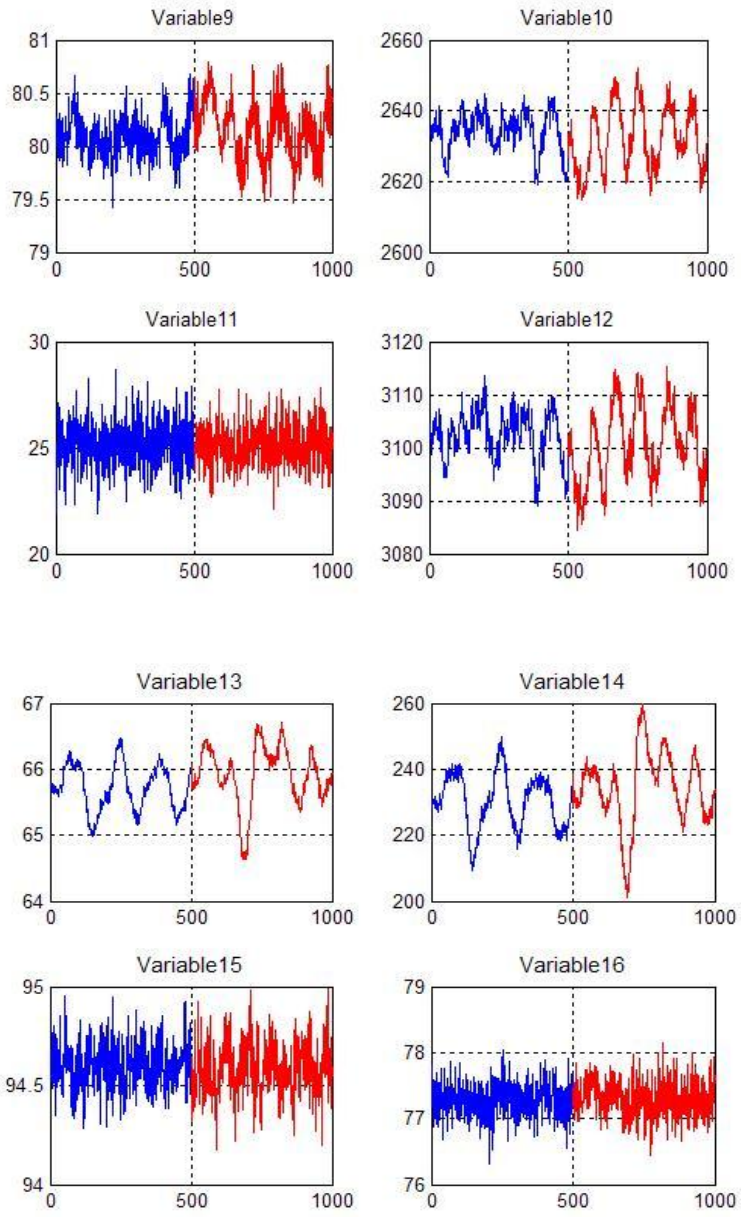


Figure 5.30: Variables (9-16) in Case 3 for normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in the TE-Process. In each plot the horizontal axis represents the number of samples and the vertical axis represents the corresponding values of each variable.

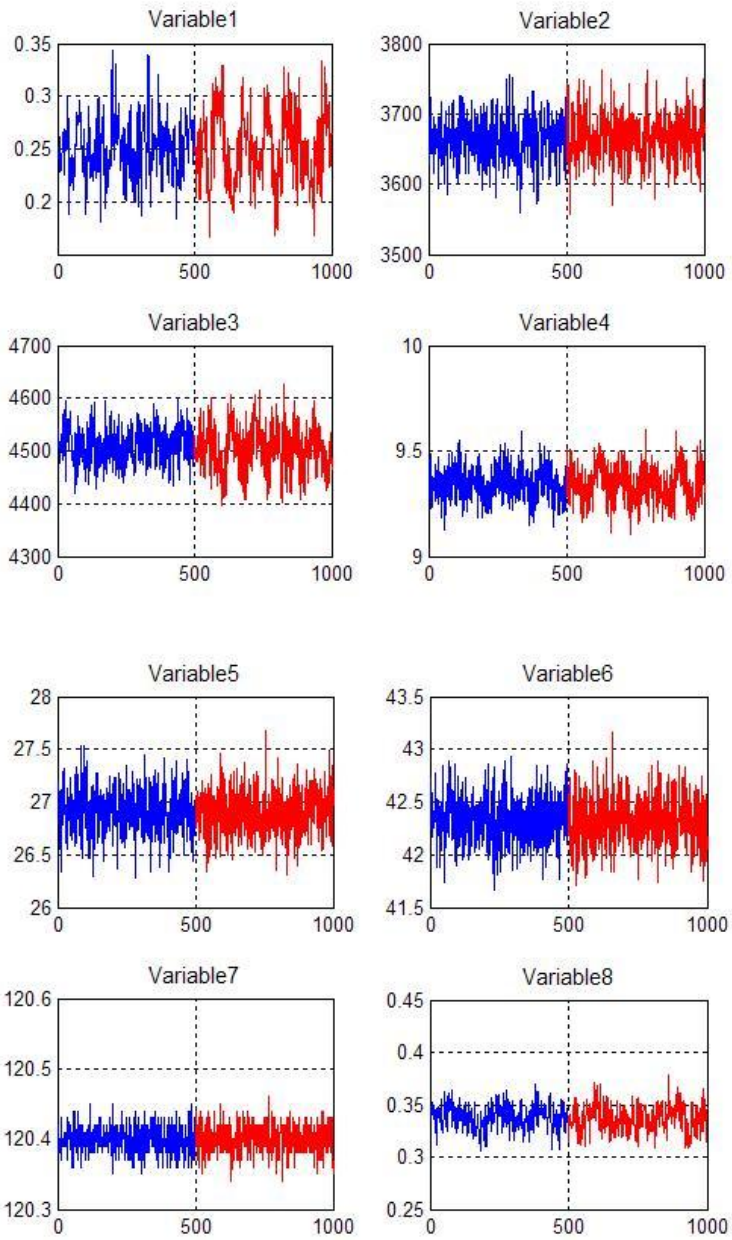


Figure 5.31: Variables (1-8) in Case 9 for normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in the TE-Process. In each plot the horizontal axis represents the sample size and the vertical axis represents the corresponding values of each variable.

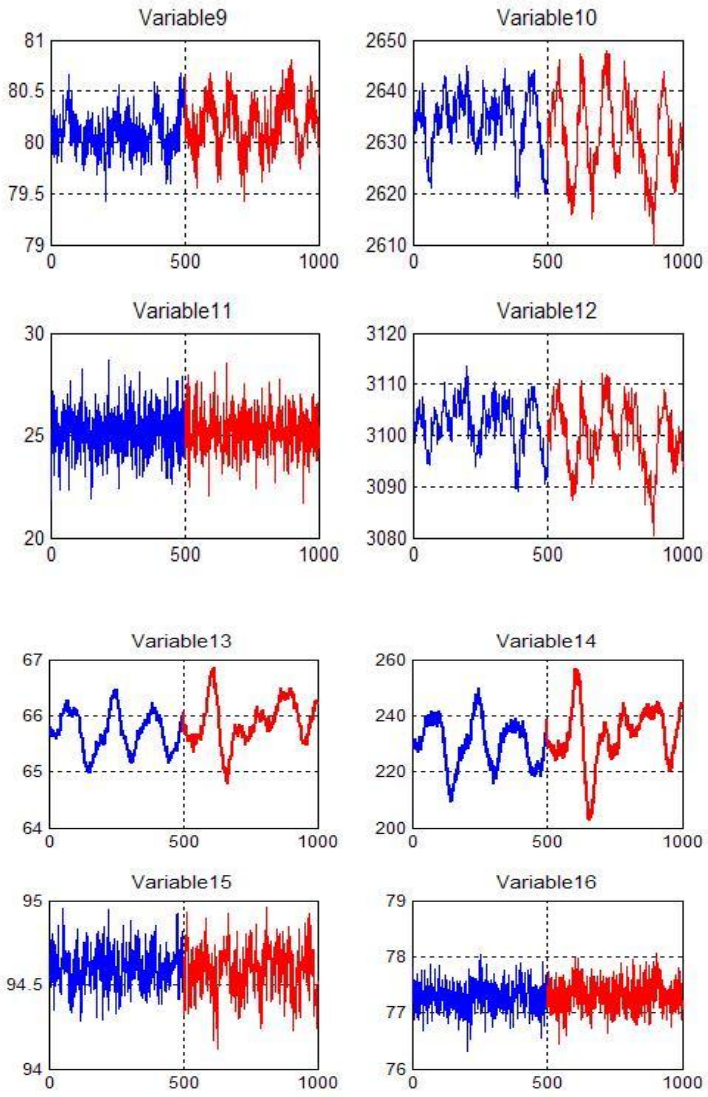


Figure 5.32: Variables (9-16) in Case 9 for normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in the TE-Process. In each plot the horizontal axis represents the number of samples and the vertical axis represents the corresponding values of each variable.

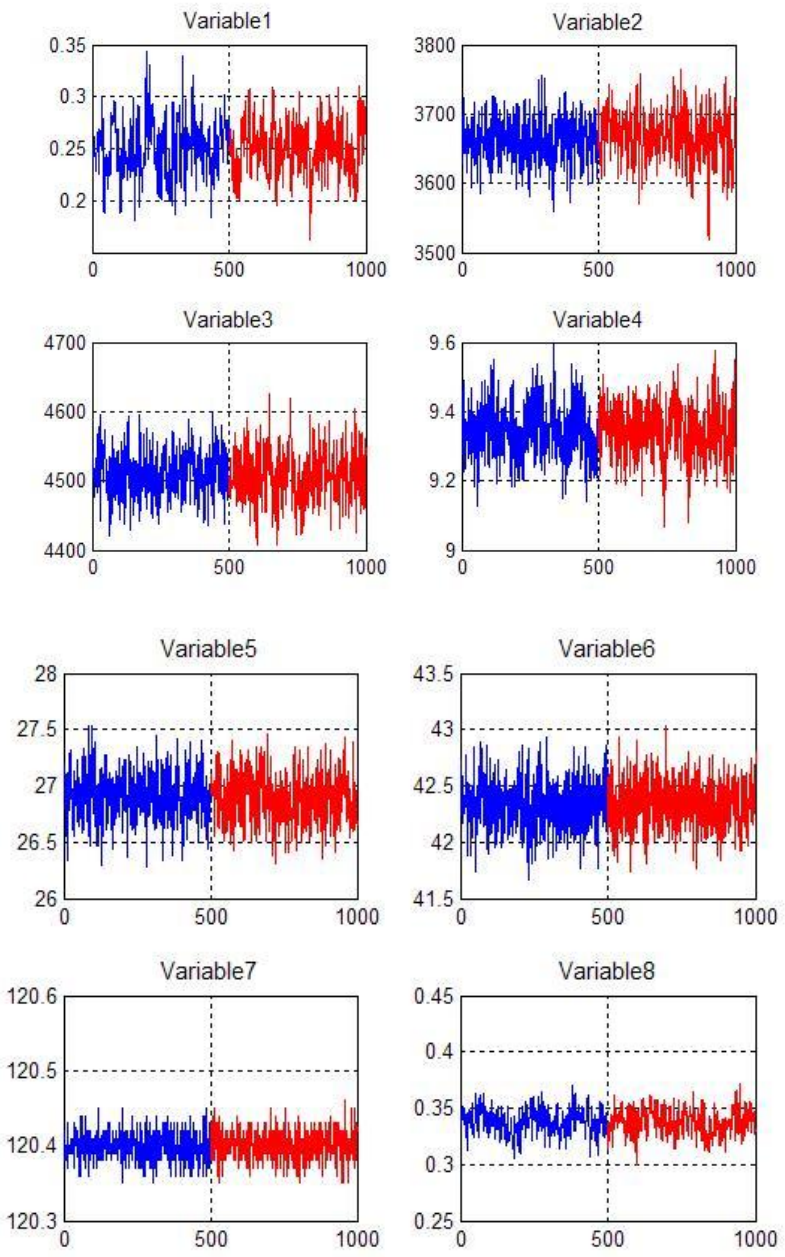


Figure 5.33: Variables (1-8) in Case 15 for normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in the TE-Process. In each plot the horizontal axis represents the sample size and the vertical axis represents the corresponding values of each variable.

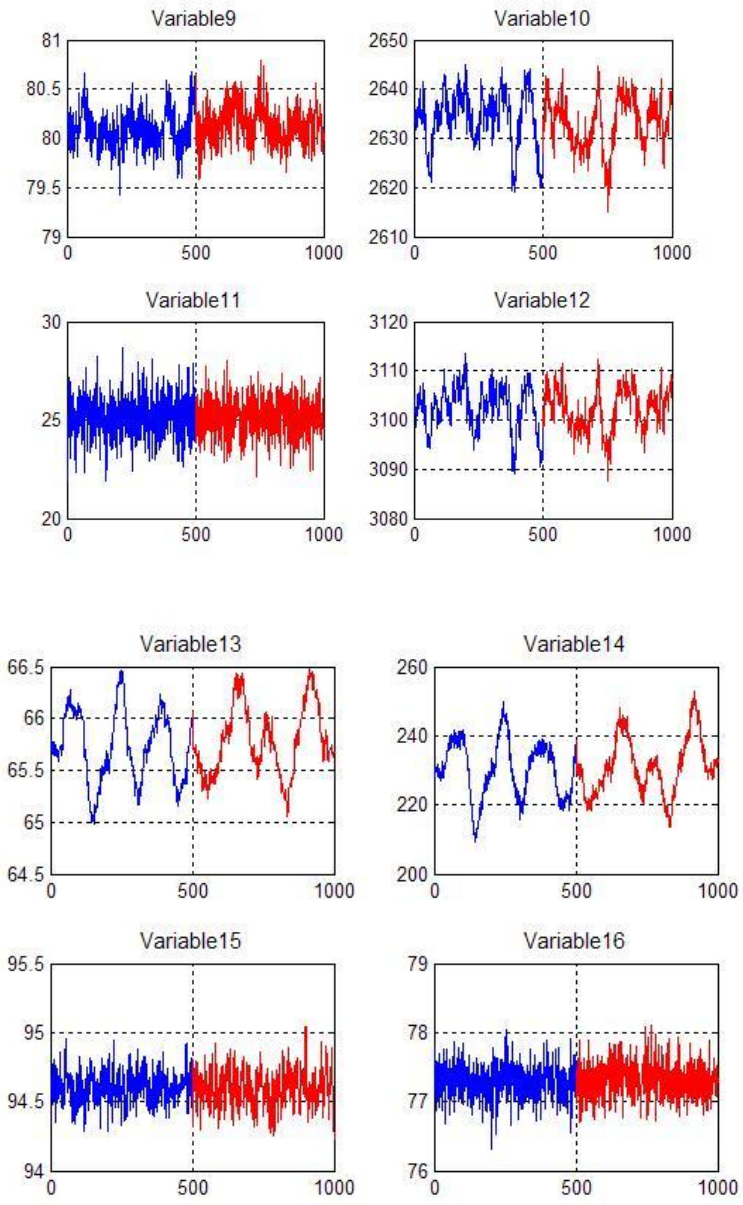
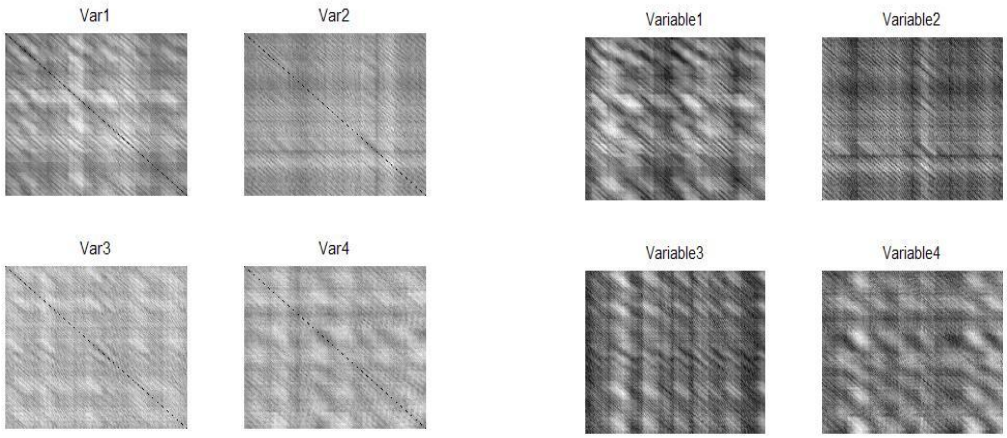


Figure 5.34: Variables (9-16) in Case 15 for normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in the TE-Process. In each plot the horizontal axis represents the number of samples and the vertical axis represents the corresponding values of each variable.

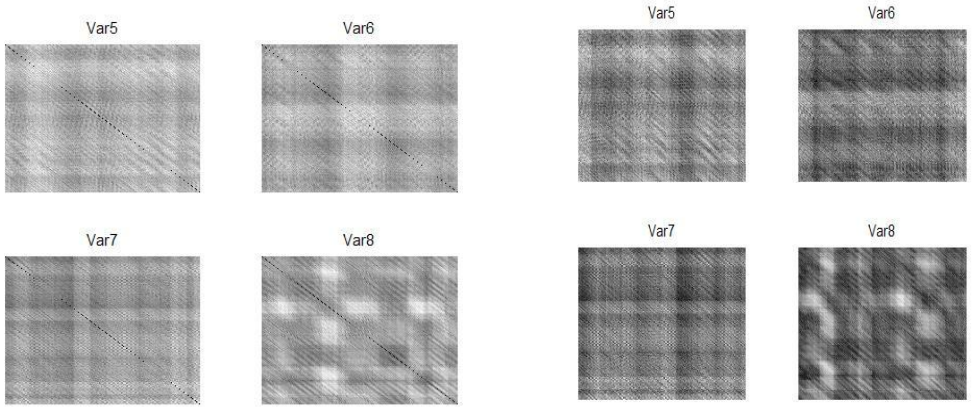
Fault 3 is the small constant change in feed concentration, which affects the steady state in the reactor. In order to keep the level of conversion of a highly exothermic reaction at a desired level, the manipulated variable, *reactor cooling water flow*, needs to be changed/controlled in such a way that any change in the mean of the steady state reactor temperature can be eliminated (Shams *et al*., 2010). Hence, any change in the

manipulated variables will have the associated change in the process variables. This type of small deviations in the process can be detected by integrating several fault detection methods as is the case in this study. Dissimilarity plots of the variables in the normal and abnormal data in this fault case can be used to visualise the global structure (similarity between different objects is reflected by their position relative to each other) of the data points in those conditions as shown in Figures 5.35–5.38. In these plots, the smaller dissimilarity values appear close to the main diagonal. Therefore, it can be visualised that the closer objects are together in these plots the higher are their similarity.



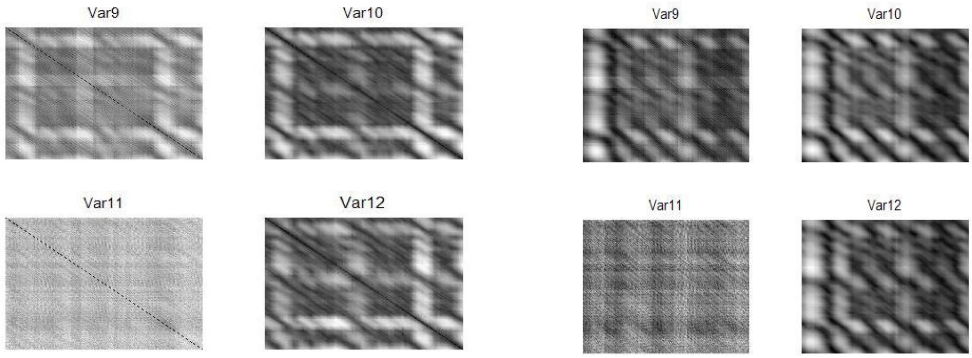
(a) (b)

Figure 5.35: (a) Dissimilarity plots of variables (1-4) in normal operating conditions (NOC) and (b) dissimilarity plots of variables (1-4) in abnormal operating conditions (ANOC) in Case 3 in TE-Process.



(a) (b)

Figure 5.36: (a) Dissimilarity plots of variables (5-8) in normal operating conditions (NOC) and (b) dissimilarity plots of variables (5-8) in abnormal operating conditions (ANOC) in Case 3 in TE-Process.



(a) (b)

Figure 5.37: (a) Dissimilarity plots of variables (9-12) in normal operating conditions (NOC) and (b) dissimilarity plots of variables (9-12) in abnormal operating condition (ANOC) in Case 3 in TE-Process.

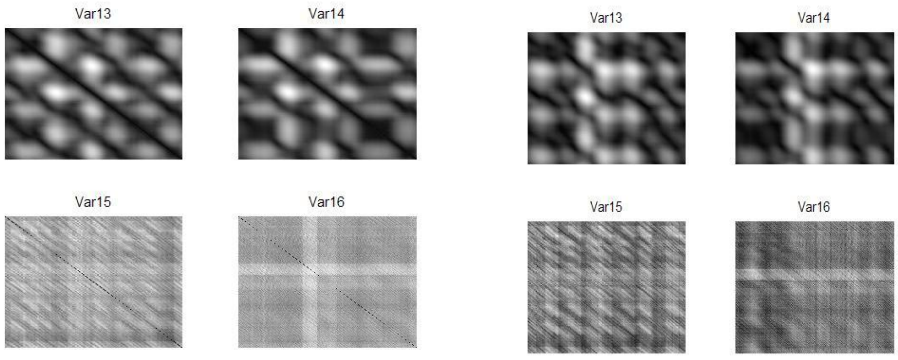


Figure 5.38: (a) Dissimilarity plots of variables (13-16) in normal operating conditions (NOC) and (b) dissimilarity plots of variables (13-16) in abnormal operating conditions (ANOC) in Case 3 in the TE-Process.

In these plots, the objects in the dissimilarity matrix are clustered together in such a way that the clusters, which contain objects that are more similar, are displayed close together. This helps to visualise the relationship between different clusters in the plot. Moreover, in the above plots the dark gray shading (or black) represents low dissimilarity values while the light gray (or white) represents high dissimilarity. The dissimilarity plots of the variables in the above Figures 5.35-5.38 are not identical and hence, clearly indicate the variations in the data structure of the variables in the normal and abnormal operating conditions. Moreover, the variations in the mean dissimilarity value of each variables in normal and abnormal conditions in Case 3 in Figure 5.39 are also compared by using the paired t-test with a significance level of 0.05. The variations in the means of the dissimilarity values for most of the variable in Figure 5.39 (except variable 6) indicate the structural difference in the variables between the normal condition and fault condition (Case 3). The p-values for the paired t-test is shown in Figure 5.40.The p- values for most of the variables in Figure 5.40 are below the significance level except for variable 6 which also indicate the structural variations in the variables in normal and faulty condition. Thus, all these results indicate the structural changes in variables hence the better detection of faults by nonlinear SSA methods in this case.

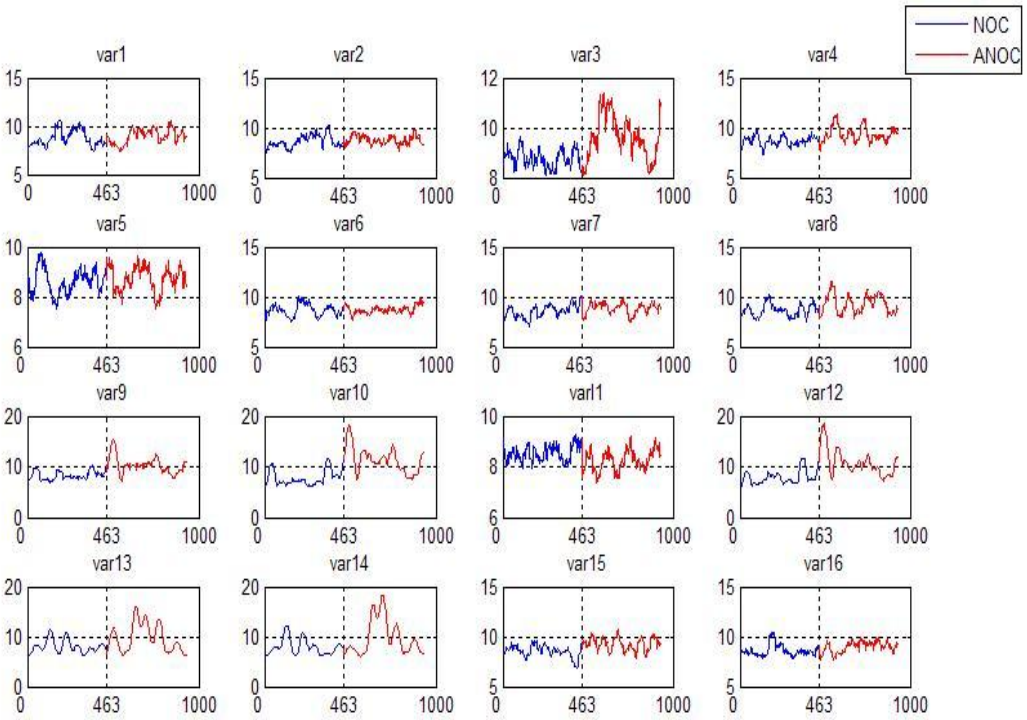


Figure 5.39: Mean value of variables in dissimilarity matrix in normal conditions (NOC) (first 463 data samples) and fault conditions (ANOC) (last 463 data samples) in Case 3. In each plot the horizontal axis represents the number of observations in the mean value of the dissimilarity matrix of the lagged variables and the vertical axis represents their corresponding mean values.

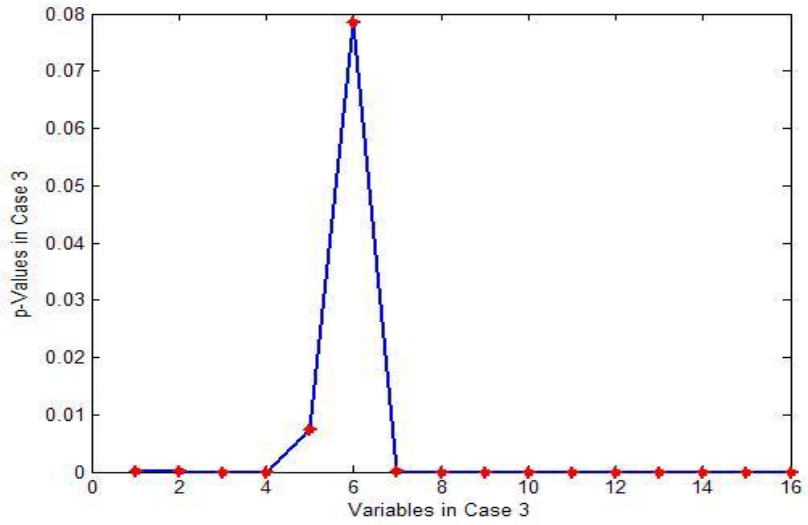


Figure 5.40: The p-values for the paired t-test on the mean values of variables in dissimilarity matrix in normal and fault condition in Case 3.

Moreover, the reconstructed variables in normal and abnormal conditions in Case 3, based on SSA methods, are shown in Figures (5.41(a-b)–5.44(a-b)). SSA-based nonlinear methods approximated faulty components in multiple modes more accurately than MS-SSA in this study as illustrated in Figures 5.41(a-b) –5.44(a-b).Thus, it obtained the highest reliability percentage in fault detection in Case 3 as shown in Table 5.5. In short, the time series pattern of the reconstructed components in Figures 5.41(a-b) – 5.44(a-b) obtained by the proposed nonlinear SSA methods are sufficiently different between the faulty and normal conditions to detect faults in the process compared to patterns of those components obtained by basic SSA in Figure 5.44(a-b). This proves the efficiency of nonlinear SSA in monitoring nonlinear processes by finding trends of different resolutions as opposed to the application of basic SSA in monitoring nonlinear processes.

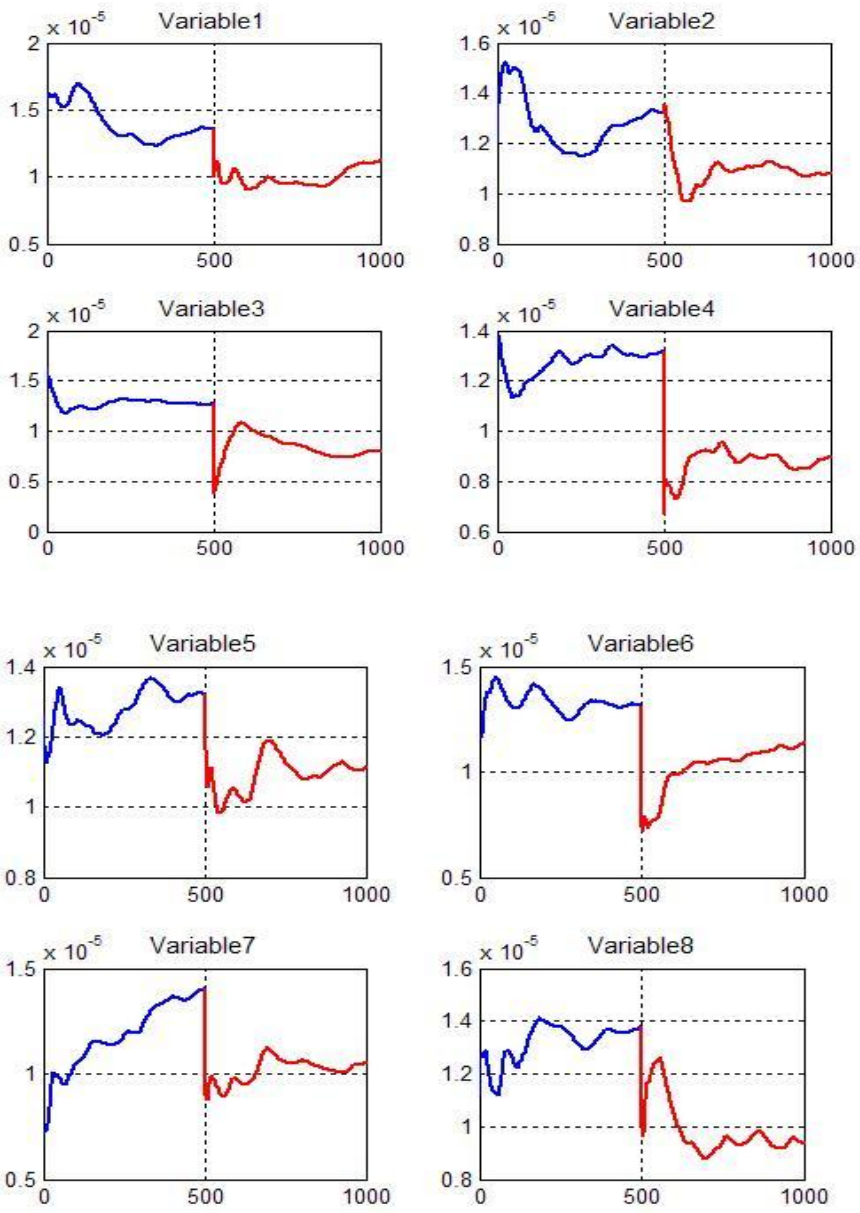


Figure 5.41 (a): The reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 1 for MM-KDSSA in case 3. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

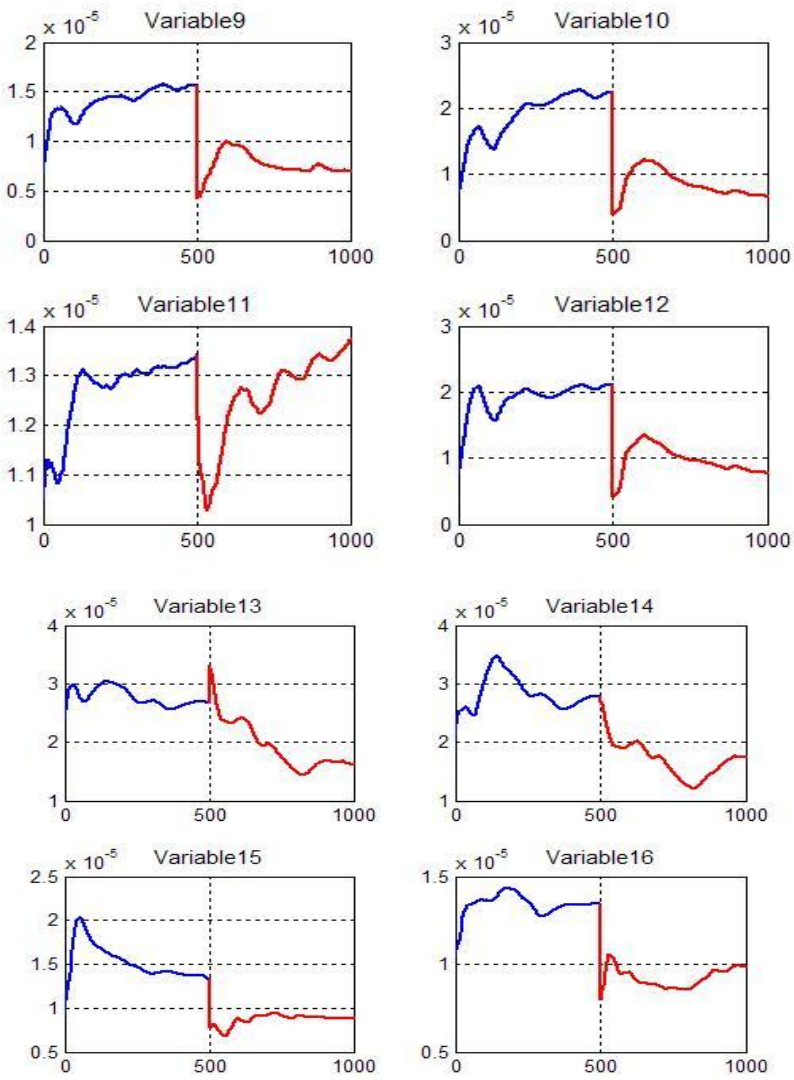


Figure 5.41(b): The reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 1 for MM-KDSSA in case 3. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

.

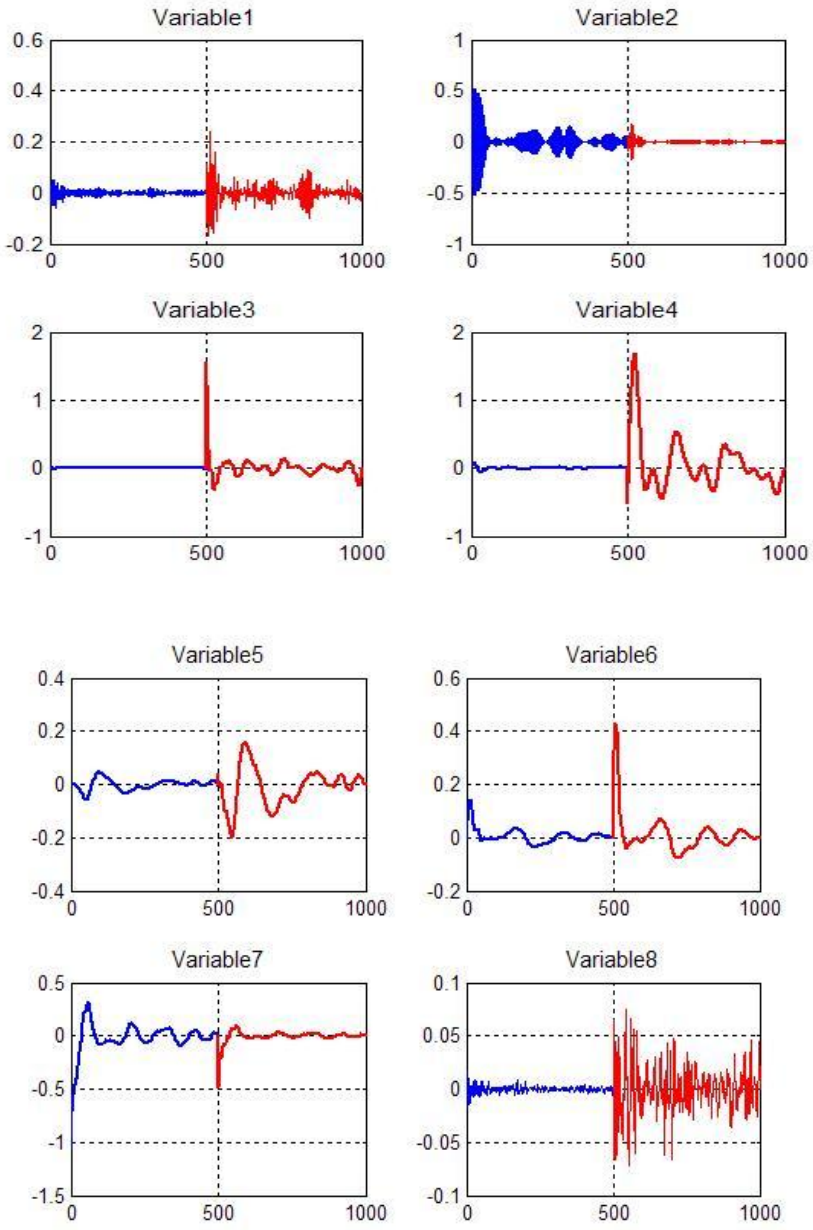


Figure 5.42 (a): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 39 for MM-DSSA in Case 3. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 39 and the vertical axis represents the corresponding values of the reconstructed variables.

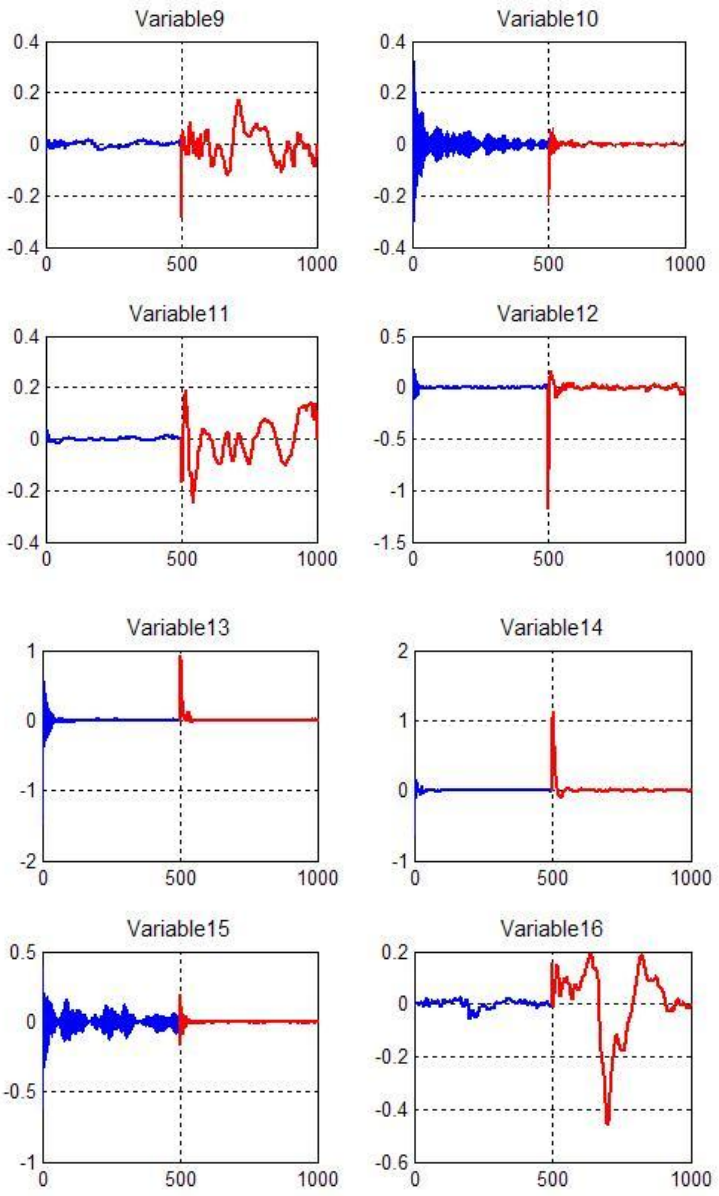


Figure 5.42(b): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 39 for MM-DSSA in Case 3. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 39 and the vertical axis represents the corresponding values of the reconstructed variables.

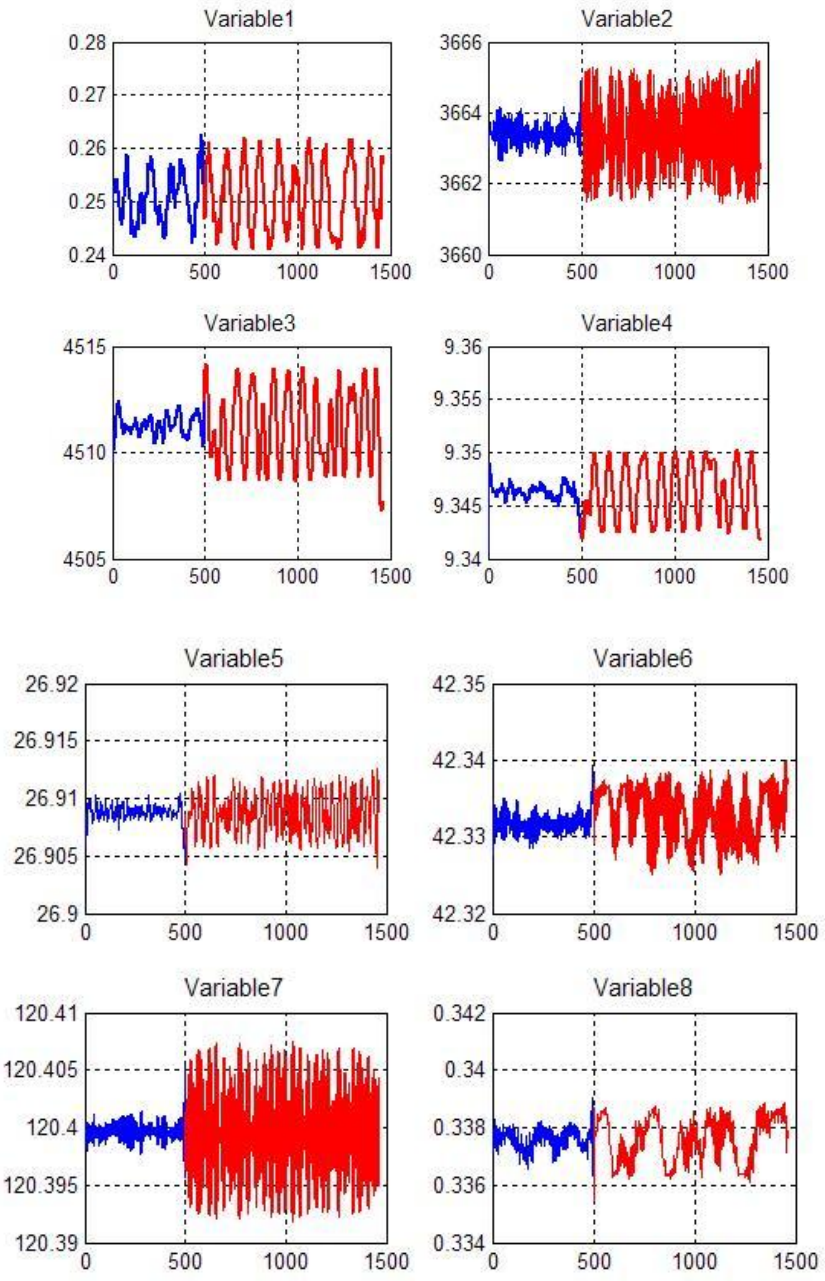


Figure 5.43 (a): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 1 for MM-NLSSA in case 3. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

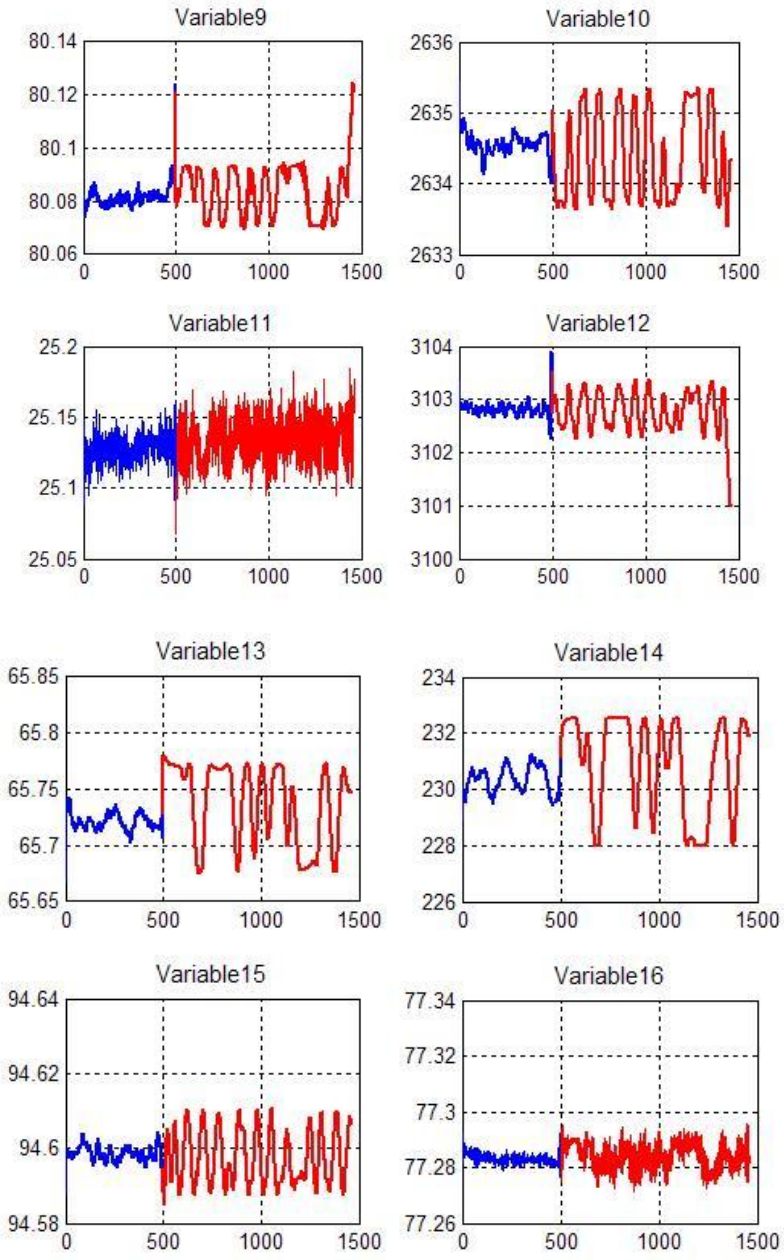


Figure 5.43(b): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 1 for MM-NLSSA in case 3. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

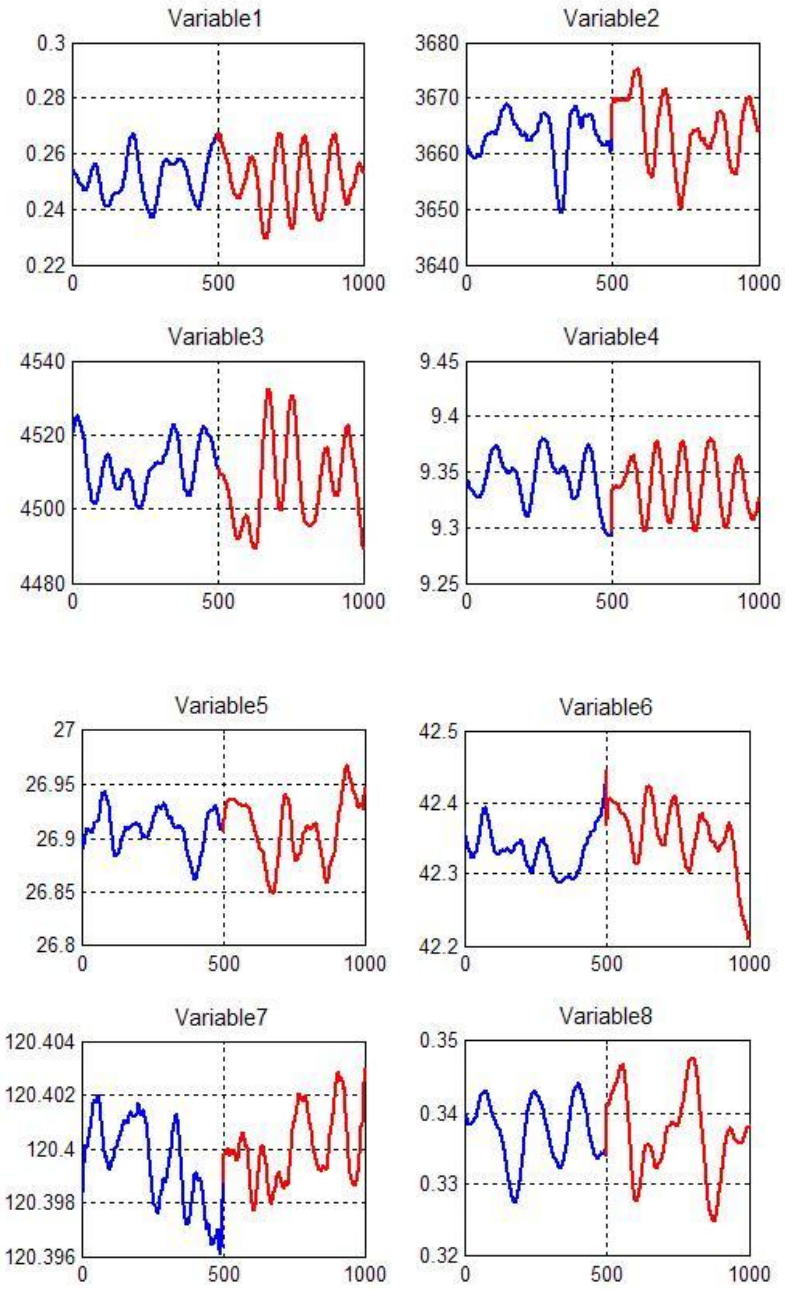


Figure 5.44 (a): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 1 for MS-SSA in case 3. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

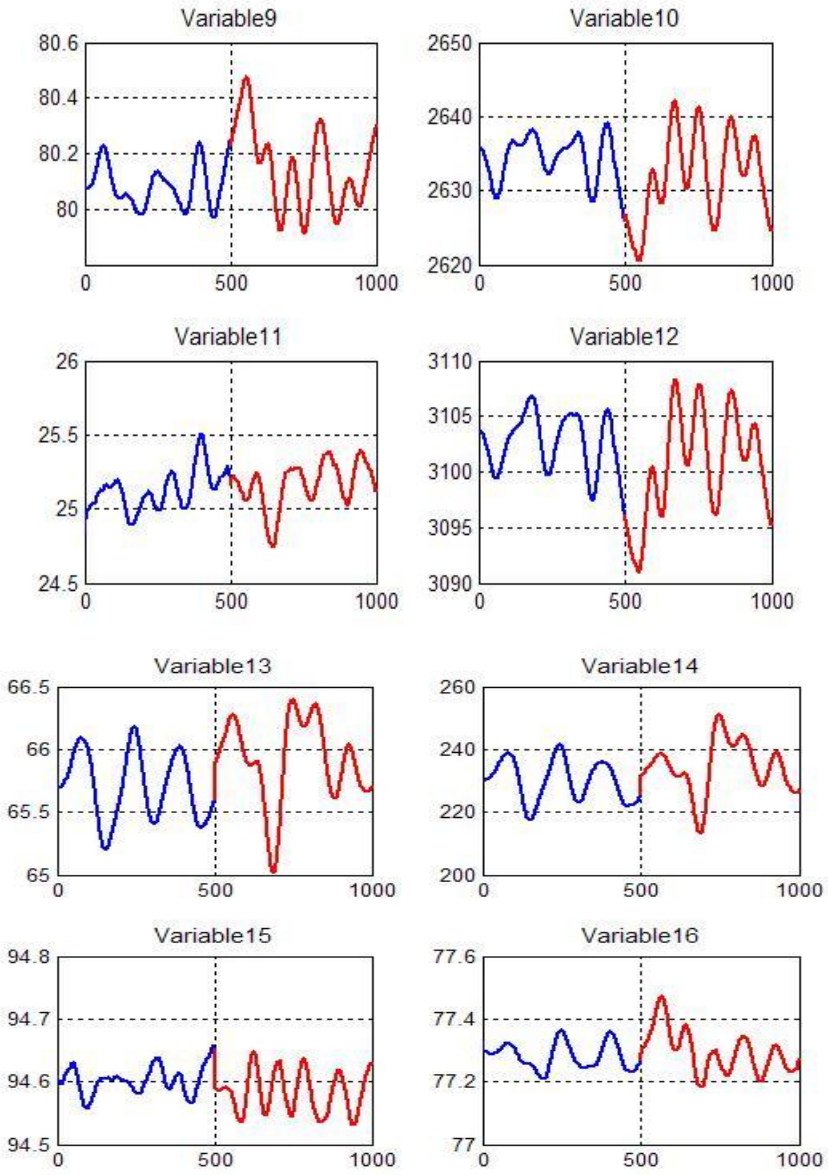
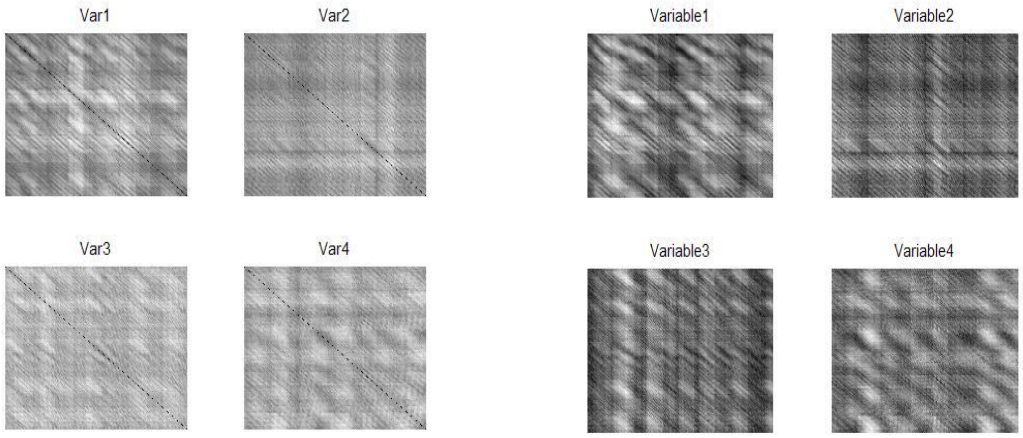


Figure 5.44(b): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 1 for MS-SSA in case 3. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

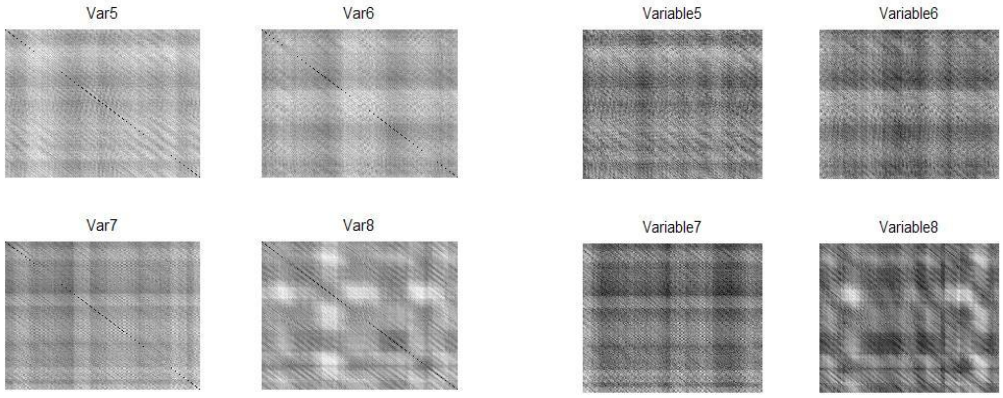
Based on the similar argument that was given in the above paragraphs for Case 3, nonlinear SSA based methods efficiently detected the changes in Case 9 and Case 15 whereas other multivariate methods and MS-SSA failed to detect faults in those cases. Case 9 is a random variation around a mean and Case 15 caused the changes in

cycling of the condenser cooling water flow due to valve stiction. Thus, these faults cause an increase in the variance of relevant manipulated variables such as reactor cooling outlet temperature and condenser cooling water flow (Shams *et al*., 2010). As mentioned earlier the faults, in Case 9 and Case 15 are also defined as unobservable from the process data and hence, it was very difficult to detect the process change in those cases with MSPC techniques and MS-SSA as shown in the previous studies (Krisnannair, 2010). The plots of the process variables in the normal and abnormal condition that was shown in Figures 5.31–5.34 also indicated that there were no observable changes in the process in these cases. Nevertheless, nonlinear SSA based process monitoring algorithms here again performed better for these unobservable faults. Random variations lead to the irregular patterns of reconstructed signals, which in turn lead to observable quantitative variations of monitoring statistics in nonlinear SSA-based algorithms. The following dissimilarity plots of variables in normal condition and abnormal conditions in Figures 5.45–5.52 and also the reconstructed components in the Figures 5.53(a-b)–5.60(a-b) obtained from nonlinear SSA decomposition and basic SSA decomposition, demonstrated these variations in the structure of the data in these fault cases and hence, better detection of faults in these cases compared to MS-SSA as shown in Table 5.5. Moreover, the mean dissimilarity values of each variable in normal and abnormal conditions in Case 9 and Case 15 are also shown in Figures 5.61–5.62. The variations in the means of the dissimilarity values for most of the variable in these cases (except for variable 2 for fault 9 and variables 8, 12, 13 and 14 for fault 15) indicated the structural difference in the variables between the normal conditions and fault conditions (Case 3 and Case 15). These variations in the mean dissimilarity value of each variable in normal and abnormal conditions in these cases were also compared by using the paired t-test with a significance level of 0.05. The p-values for the paired t-test are shown in Figure 5.63.The p- values for most of the variables for Case 9 and Case 15 in Figure 5.63 are below the significance level except for variable 2 in Case 9 and for variables 8, 12, 13 and 14 in Case 15. This shows the structural variations in the variables in normal and faulty conditions in Case 9 and Case 15 and hence, the better detection of faults by nonlinear SSA based methods in these cases. In short, nonlinear SSA-based methods outperformed MS-SSA and PCA in monitoring the TE-Process in most of the difficult fault conditions in this study.



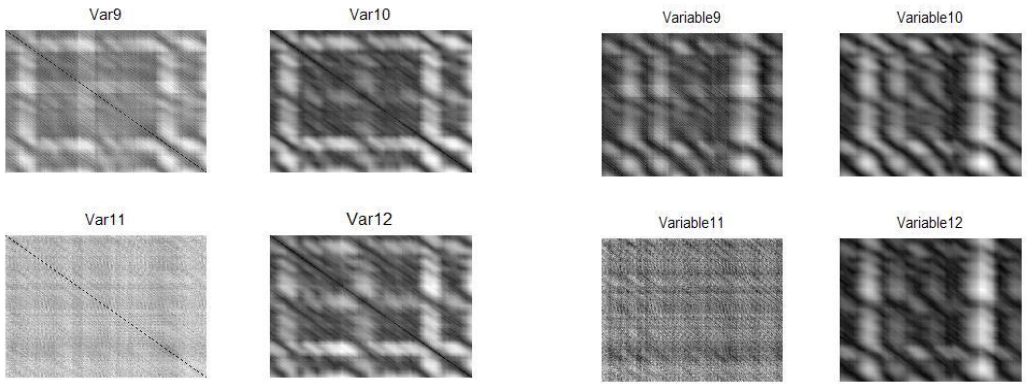
(a) (b)

Figure 5.45: (a) Dissimilarity plots of variables (1-4) in normal operating conditions (NOC) and (b) dissimilarity plots of variables (1-4) in abnormal operating conditions (ANOC) in Case 9 in TE-Process.



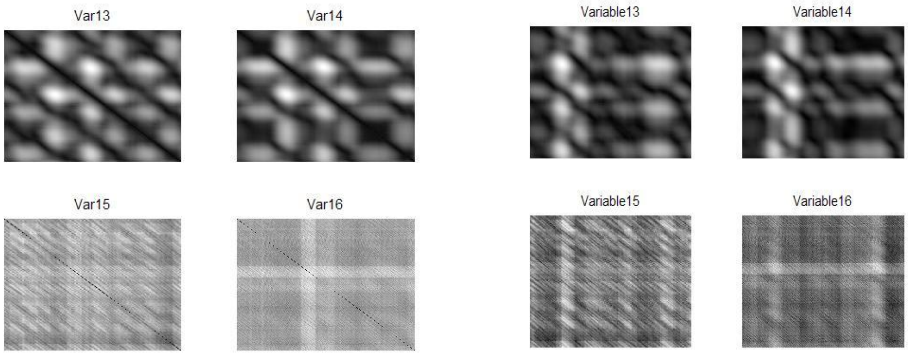
(a) (b)

Figure 5.46: (a) Dissimilarity plots of variables (5-6) in normal operating conditions (NOC) and (b) dissimilarity plots of variables (5-6) in abnormal operating conditions (ANOC) in Case 9 in TE-Process.



(a) (b)

Figure 5.47: (a) Dissimilarity plots of variables (9-12) in normal operating conditions (NOC) and (b) dissimilarity plots of variables (9-12) in abnormal operating conditions (ANOC) in Case 9 in the TE-Process.



(a) (b)

Figure 5.48: (a) Dissimilarity plots of variables (13-16) in normal operating conditions (NOC) and (b) dissimilarity plots of variables (13-16) in abnormal operating conditions (ANOC) in Case 9 in the TE-Process.

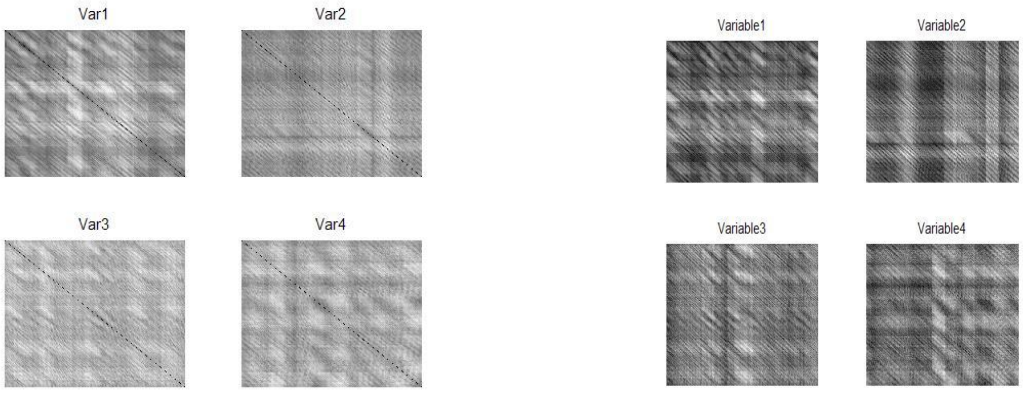


Figure 5.49: (a) Dissimilarity plots of variables (1-4) in normal operating conditions (NOC) and (b) dissimilarity plots of variables (1-4) in abnormal operating conditions (ANOC) in Case 15 in the TE-Process.

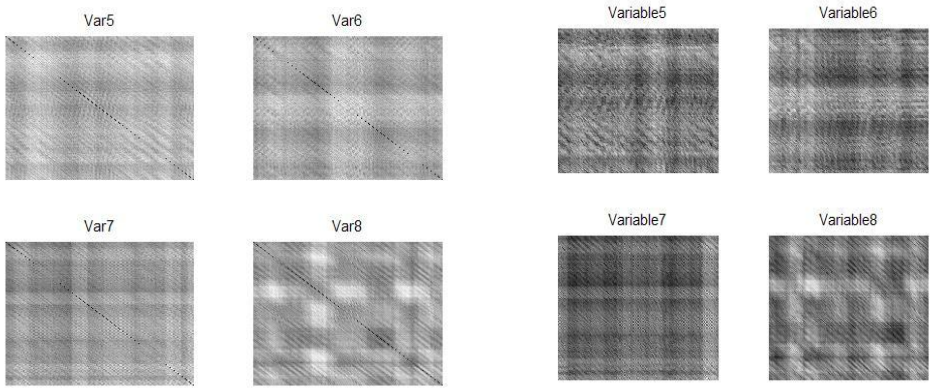


Figure 5.50: (a) Dissimilarity plots of variables (5-8) in normal operating conditions (NOC) and (b) dissimilarity plots of variables (5-8) in abnormal operating conditions (ANOC) in Case 15 in the -Process.

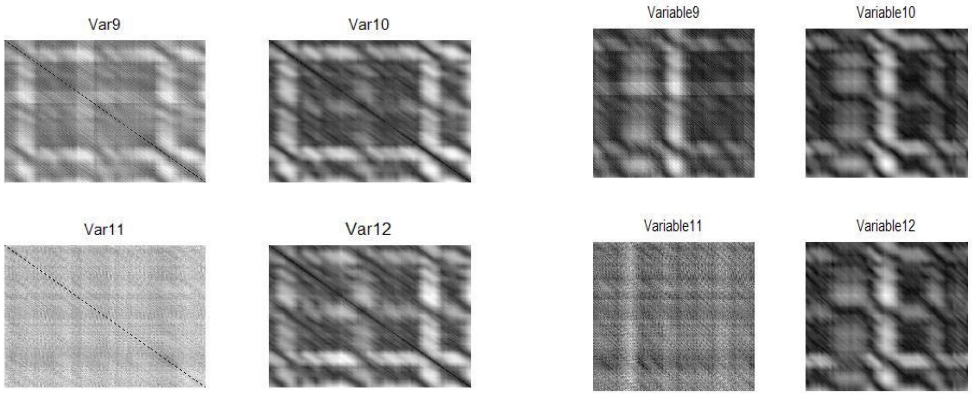


Figure 5.51: (a) Dissimilarity plots of variables (9-12) in normal operating conditions (NOC) and (b) dissimilarity plots of variables (9-12) in abnormal operating conditions (ANOC) in Case 15 in the TE-Process.

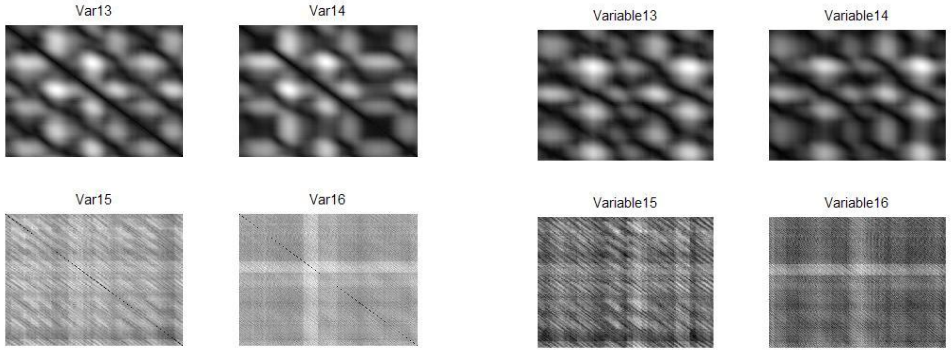


Figure 5.52: (a) Dissimilarity plots of variables (13-16) in normal operating conditions (NOC) and (b) dissimilarity plots of variables (13-16) in abnormal operating conditions (ANOC) in Case 15 in the TE-Process.

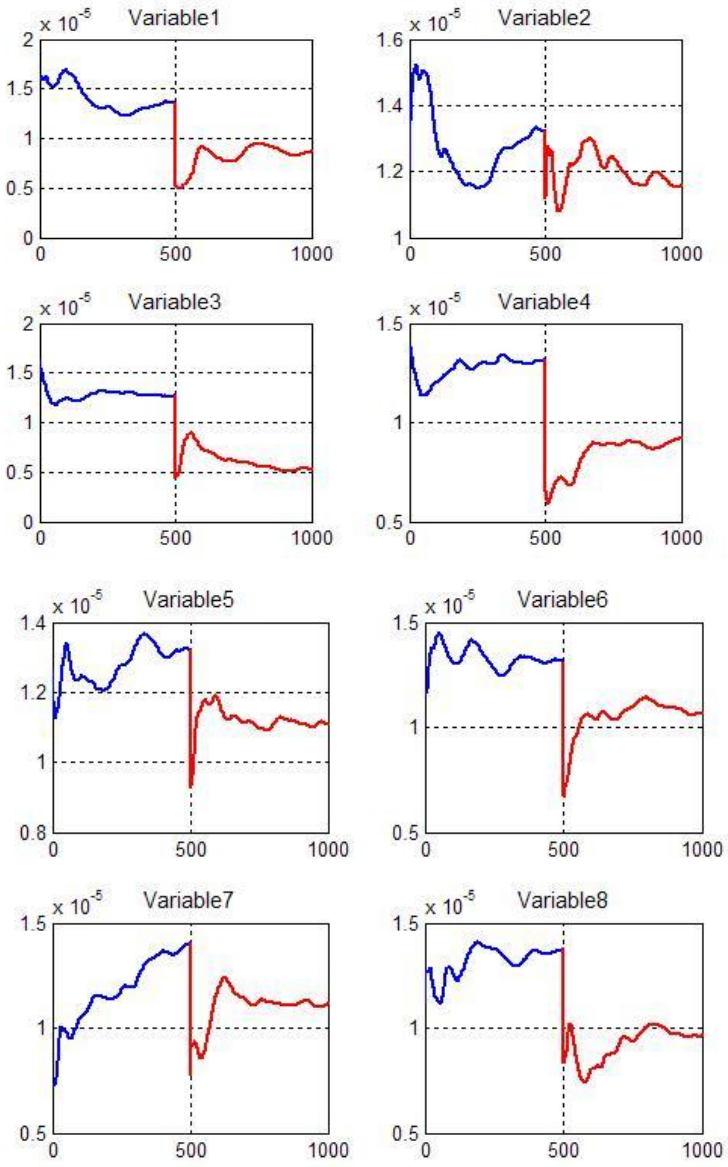


Figure 5.53 (a): Reconstructed variable in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode1 for MM-KDSSA in Case 9. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

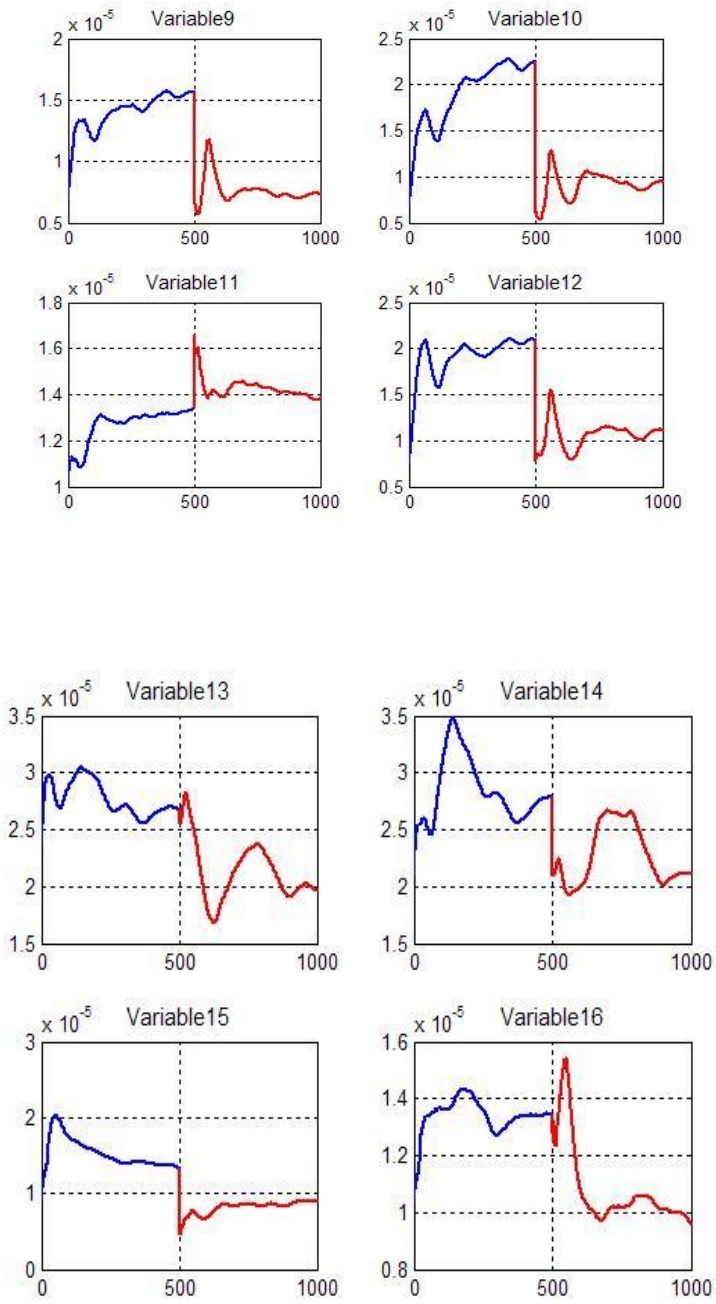


Figure 5.53(b): Reconstructed variable in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode1 for MM-KDSSA in Case 9. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

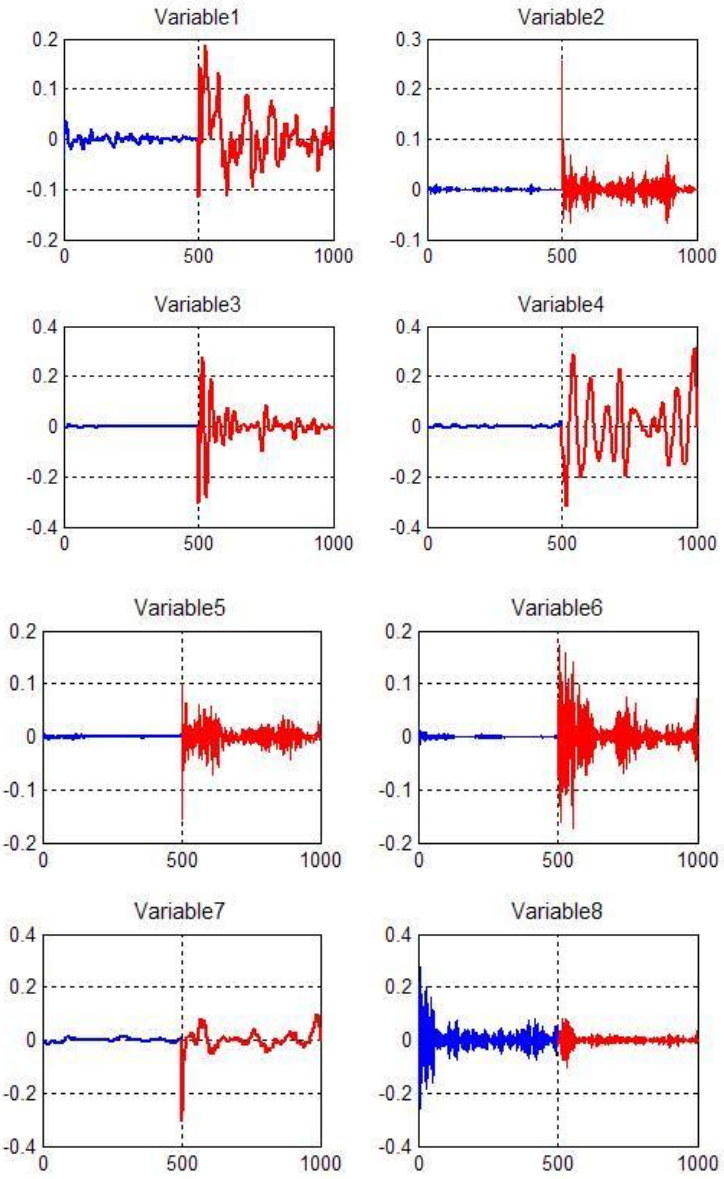


Figure 5.54(a): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 40 for MM-DSSA in case 9. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 40 and the vertical axis represents the corresponding values of the reconstructed variables.

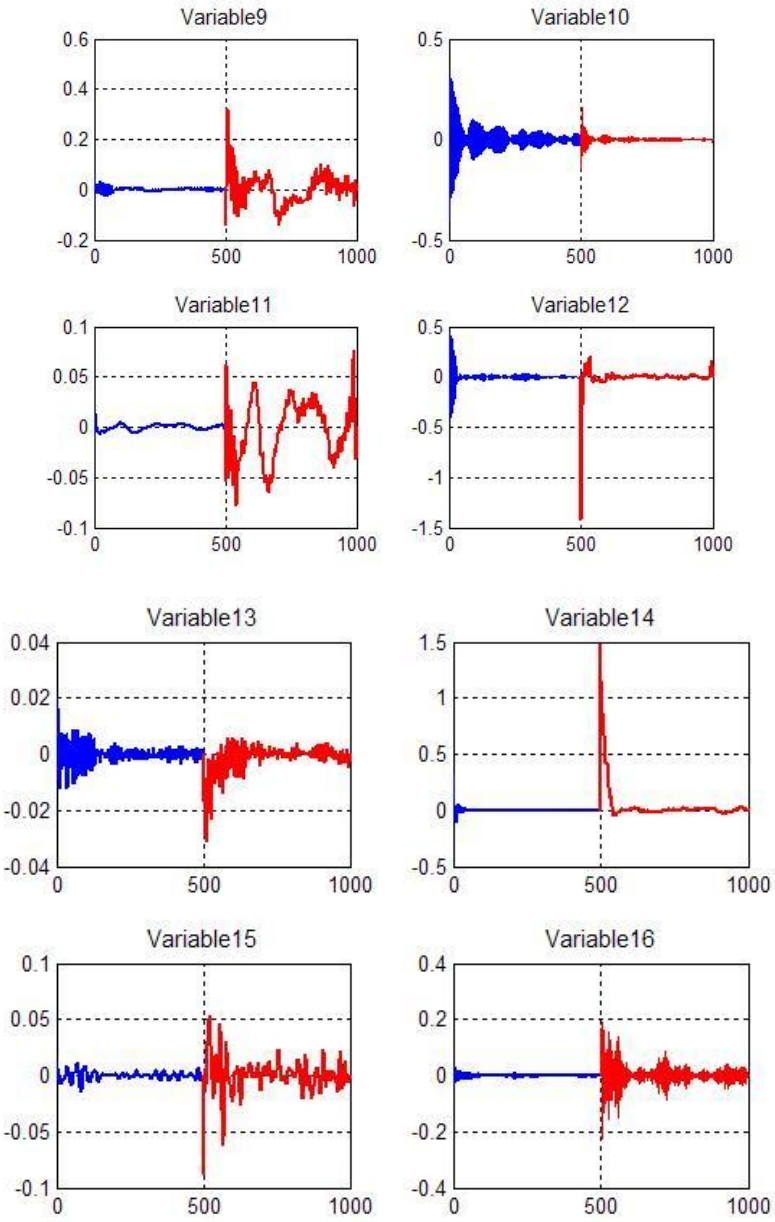


Figure 5.54(b): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 40 for MM-DSSA in case 9. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 40 and the vertical axis represents the corresponding values of the reconstructed variables.

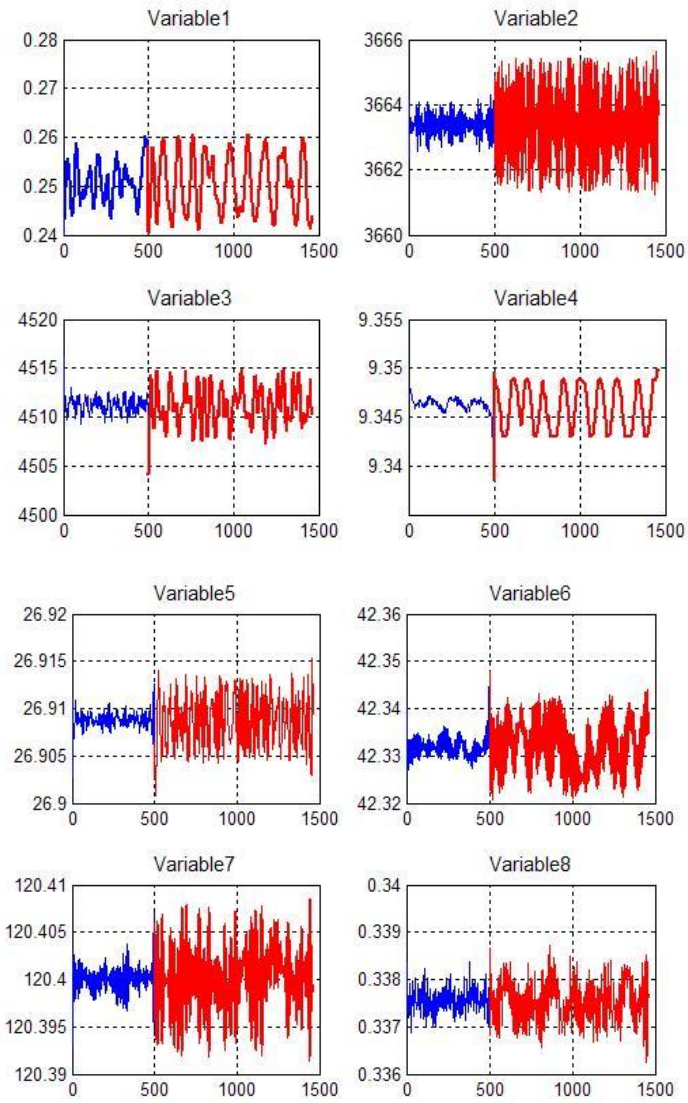


Figure 5.55 (a): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 1 for MM-NLSSA in case 9. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

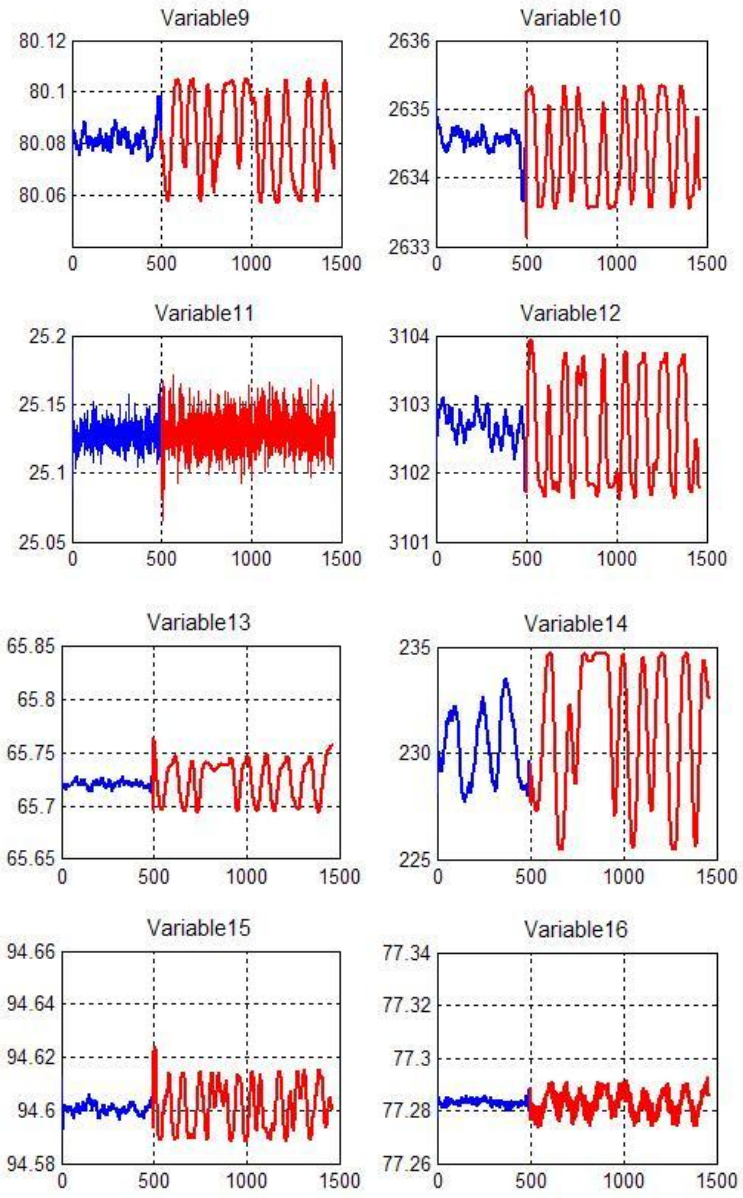


Figure 5.55 (b): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 1 for MM-NLSSA in case 9. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

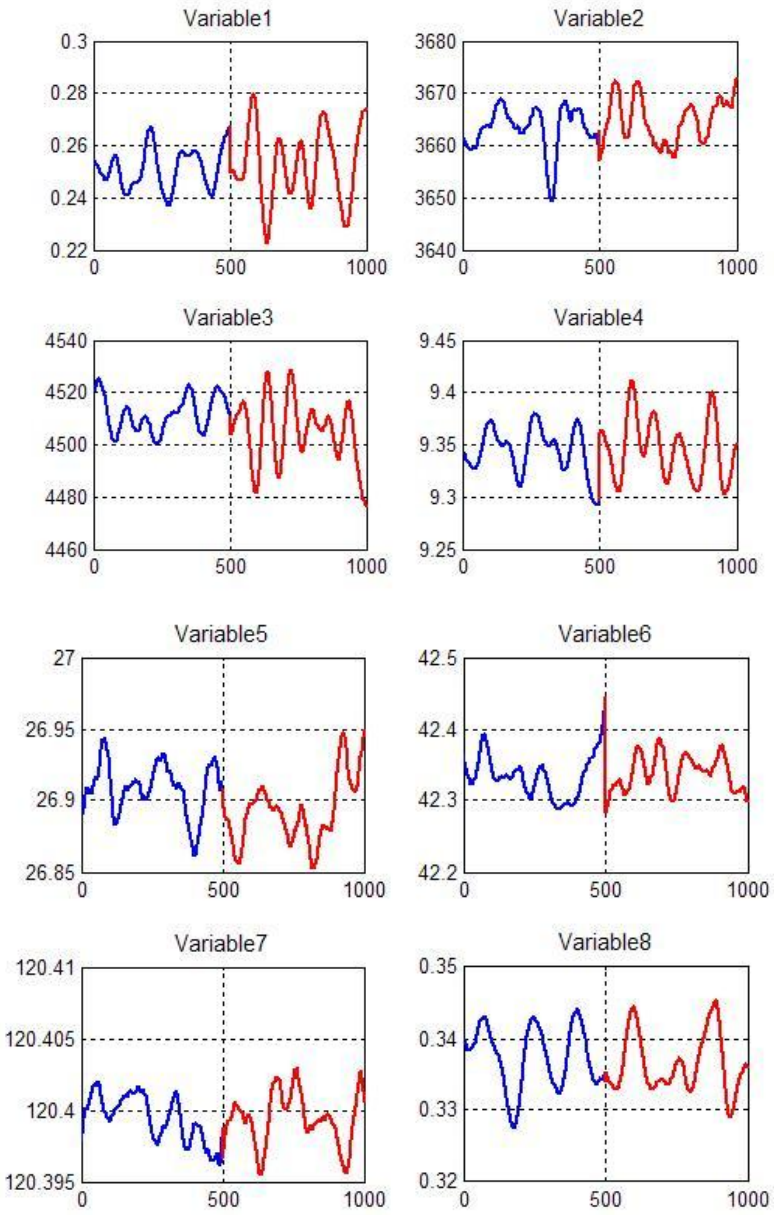


Figure 5.56 (a): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode1 for MS-SSA in case 9. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

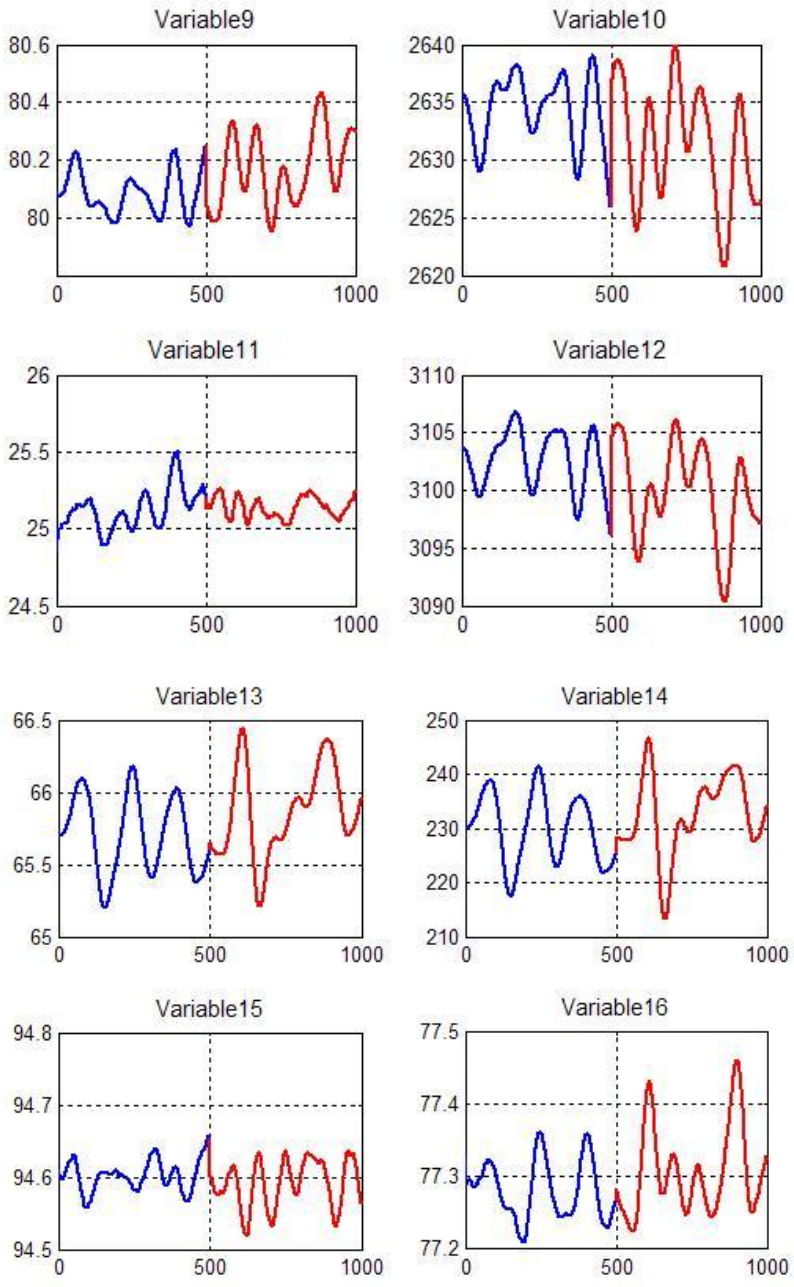


Figure 5.56 (b): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode1 for MS-SSA in case 9. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

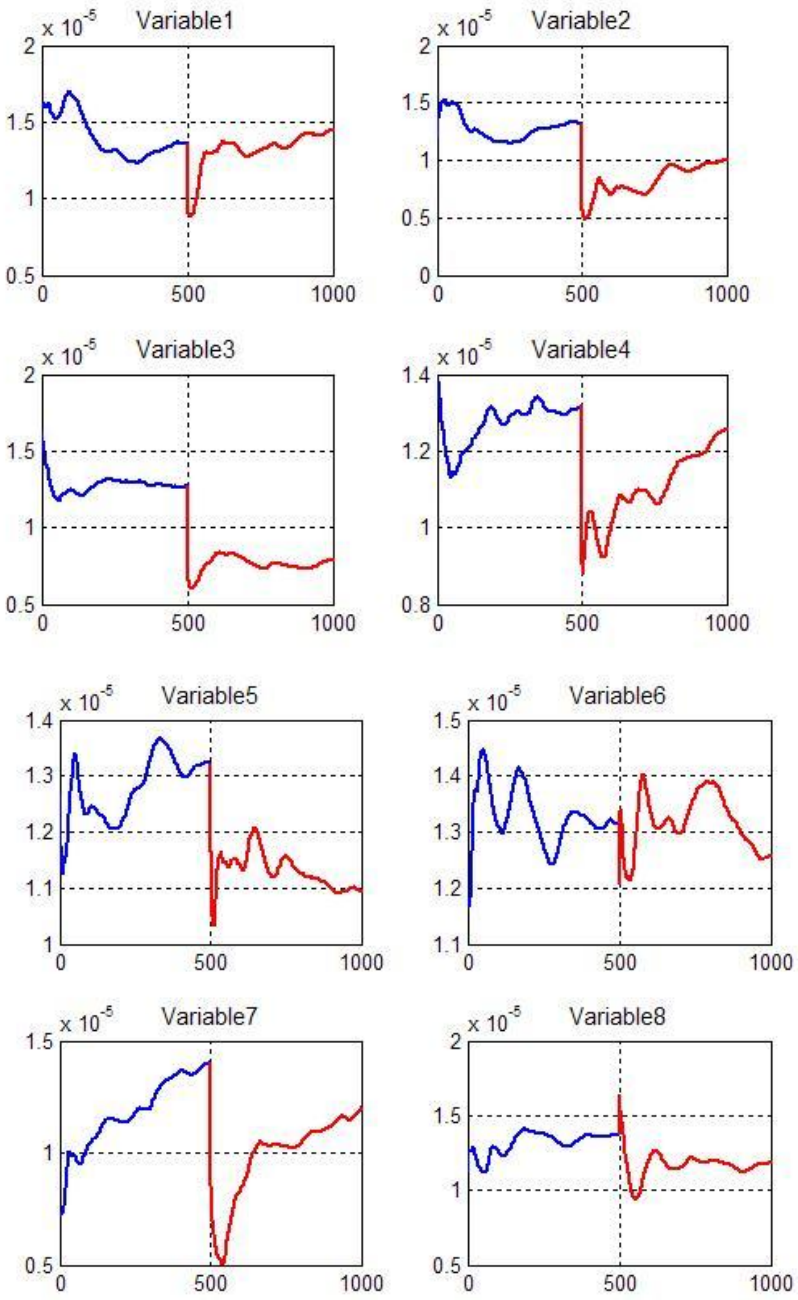


Figure 5.57 (a): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode1 for MM-KDSSA in case 15. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

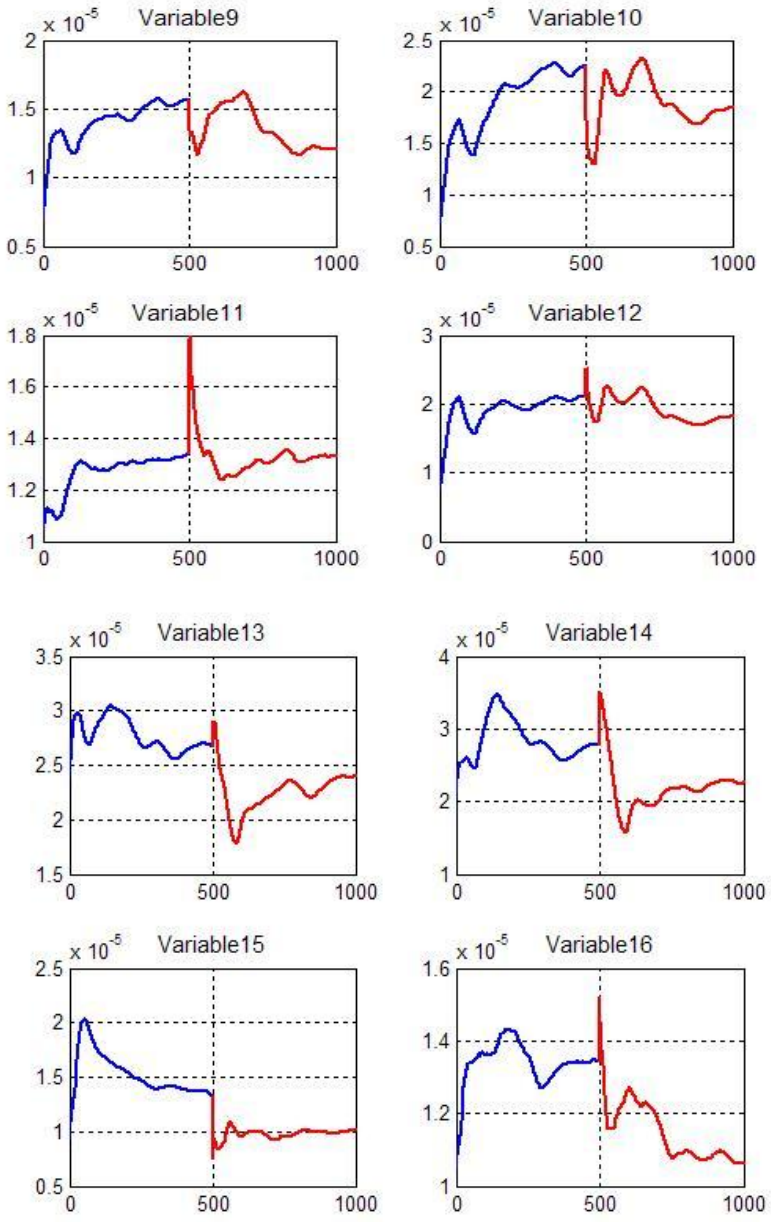


Figure 5.57(b): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode1 for MM-KDSSA in case 15. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

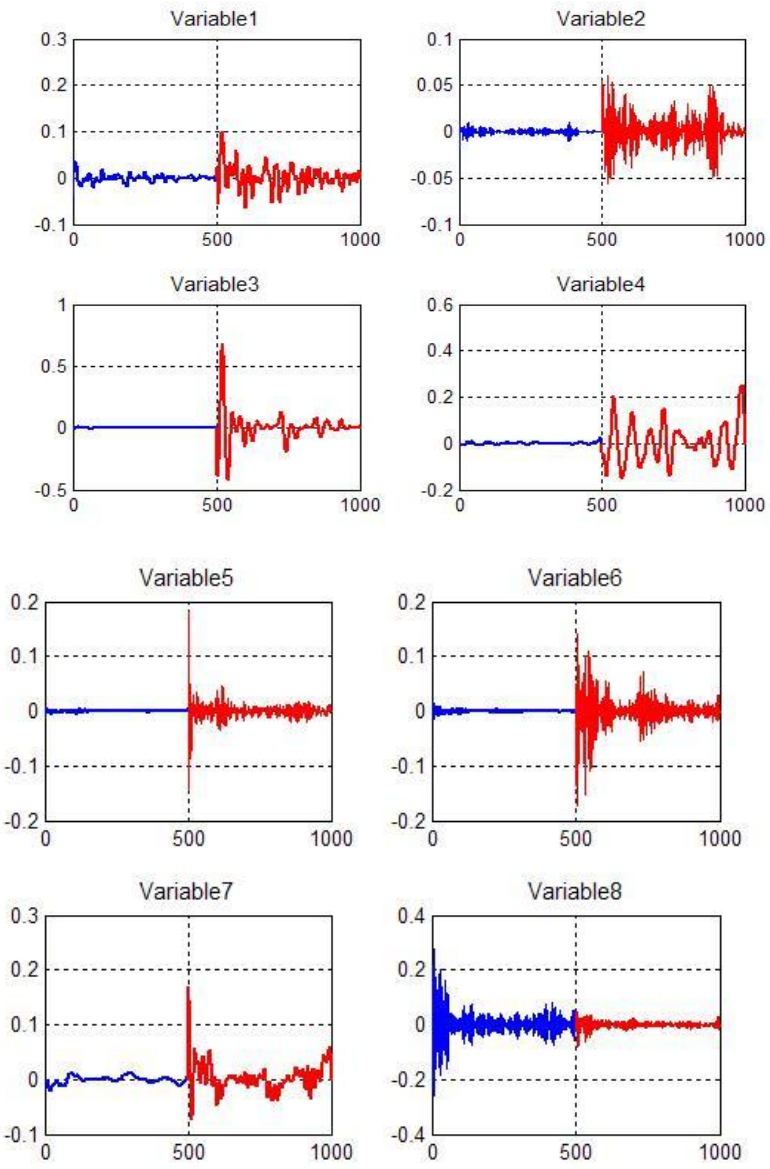


Figure 5.58 (a): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 40 for MM-DSSA in case 15. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 40 and the vertical axis represents the corresponding values of the reconstructed variables.

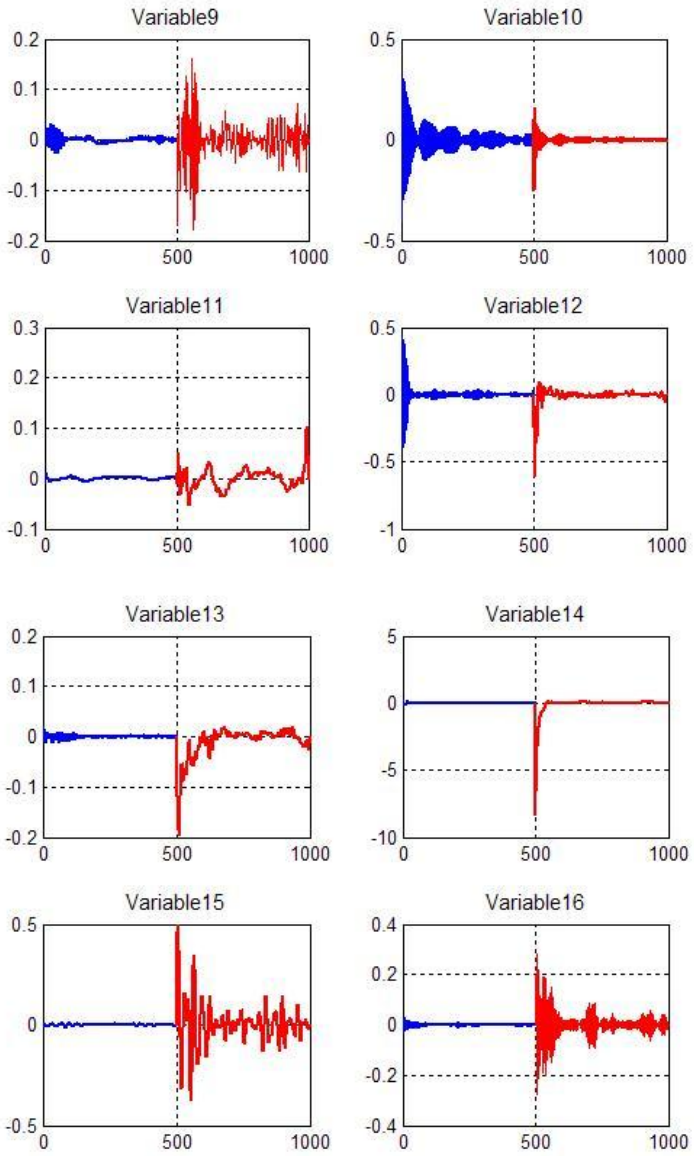


Figure 5.58(b): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 40 for MM-DSSA in case 15. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 40 and the vertical axis represents the corresponding values of the reconstructed variables.

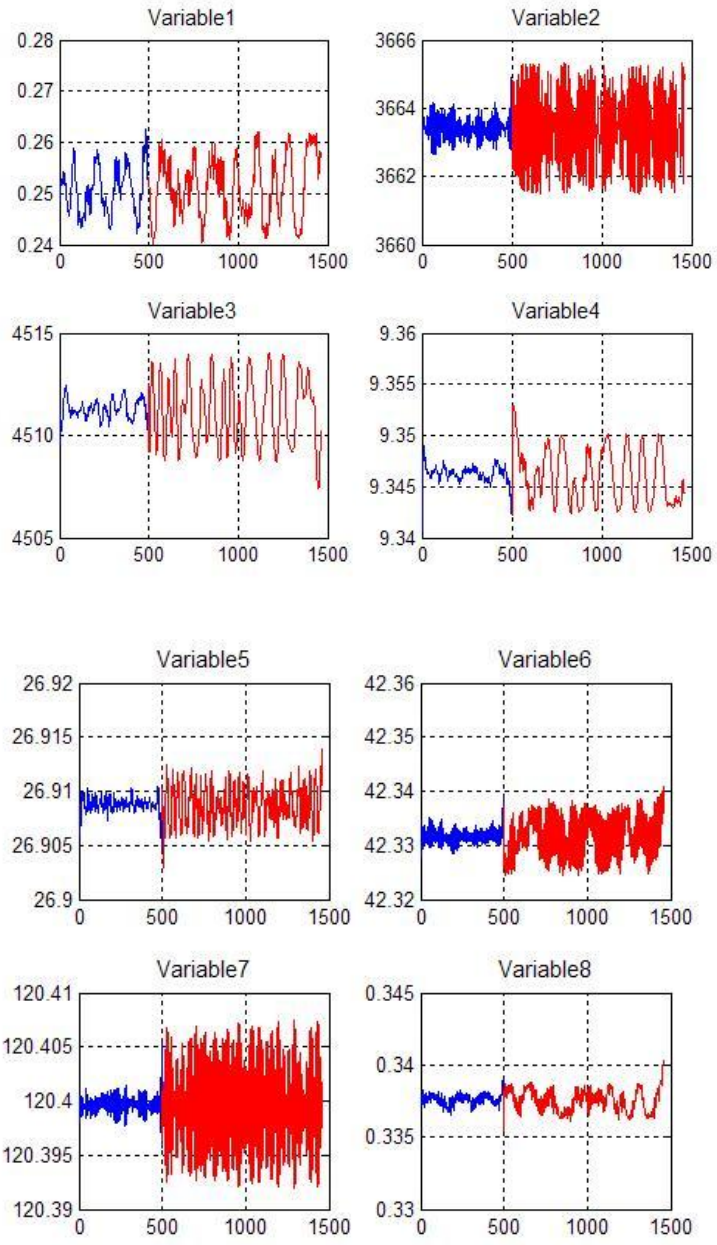


Figure 5.59 (a): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 1 for MM-NLSSA in case 15. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

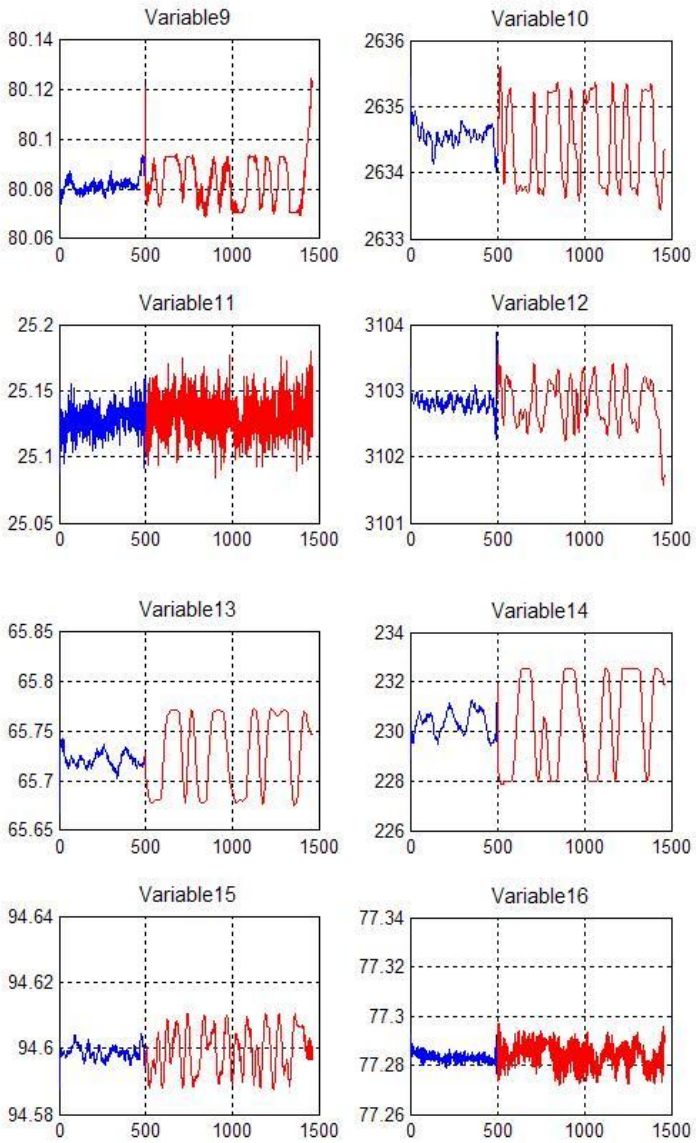


Figure 5.59(b): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 1 for MM-NLSSA in case 15. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

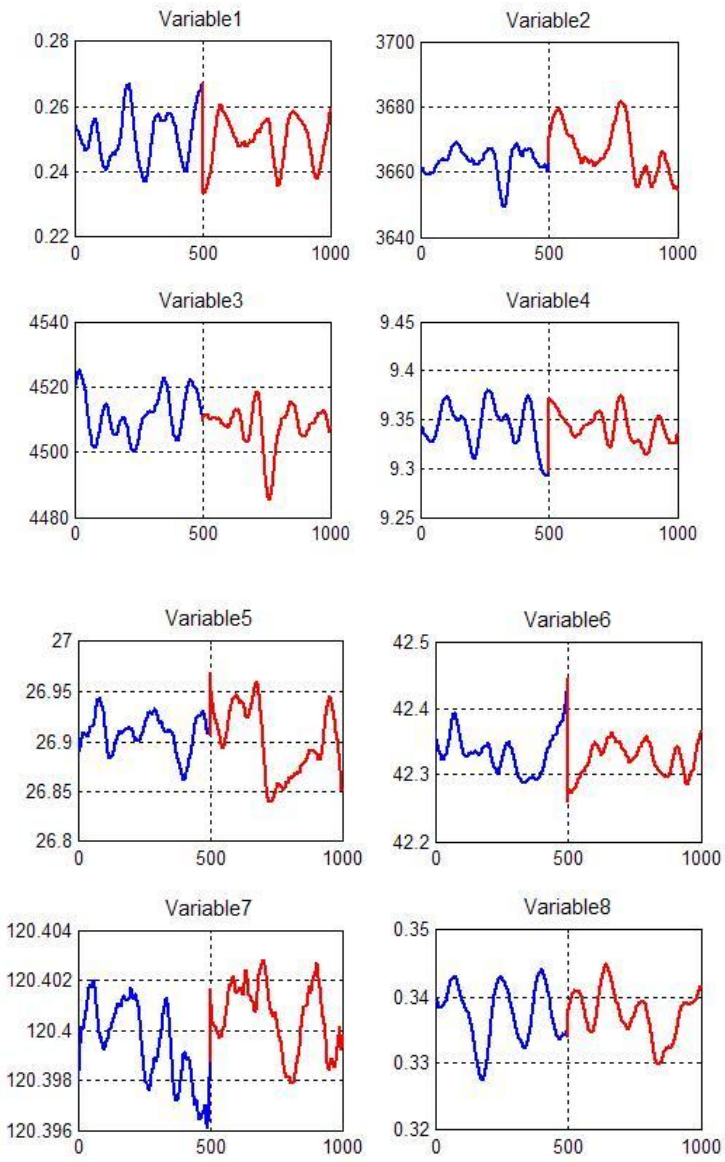


Figure 5.60 (a): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 1 for MS-SSA in case 15. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

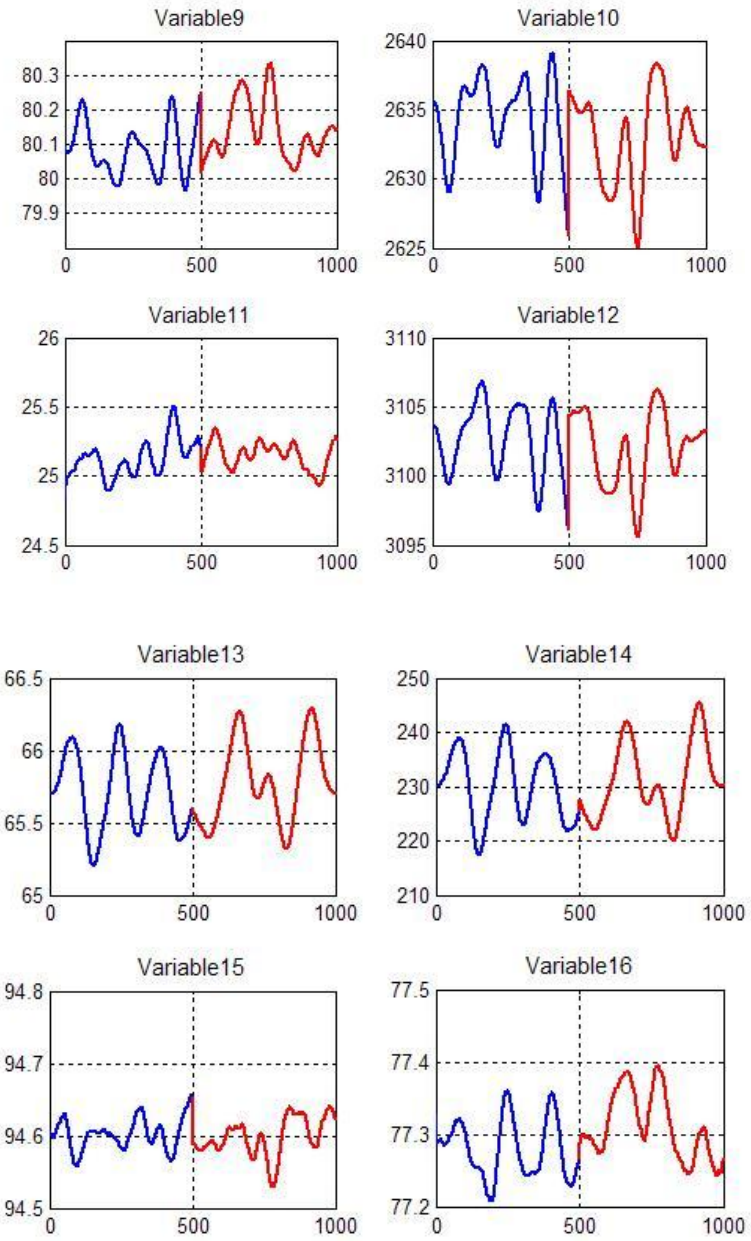


Figure 5.60(b): Reconstructed variables in normal conditions (first 500 samples) and abnormal conditions (last 500 samples) in mode 1 for MS-SSA in case 15. In each plot the horizontal axis represents the number of samples in the reconstructed components of each variable in mode 1 and the vertical axis represents the corresponding values of the reconstructed variables.

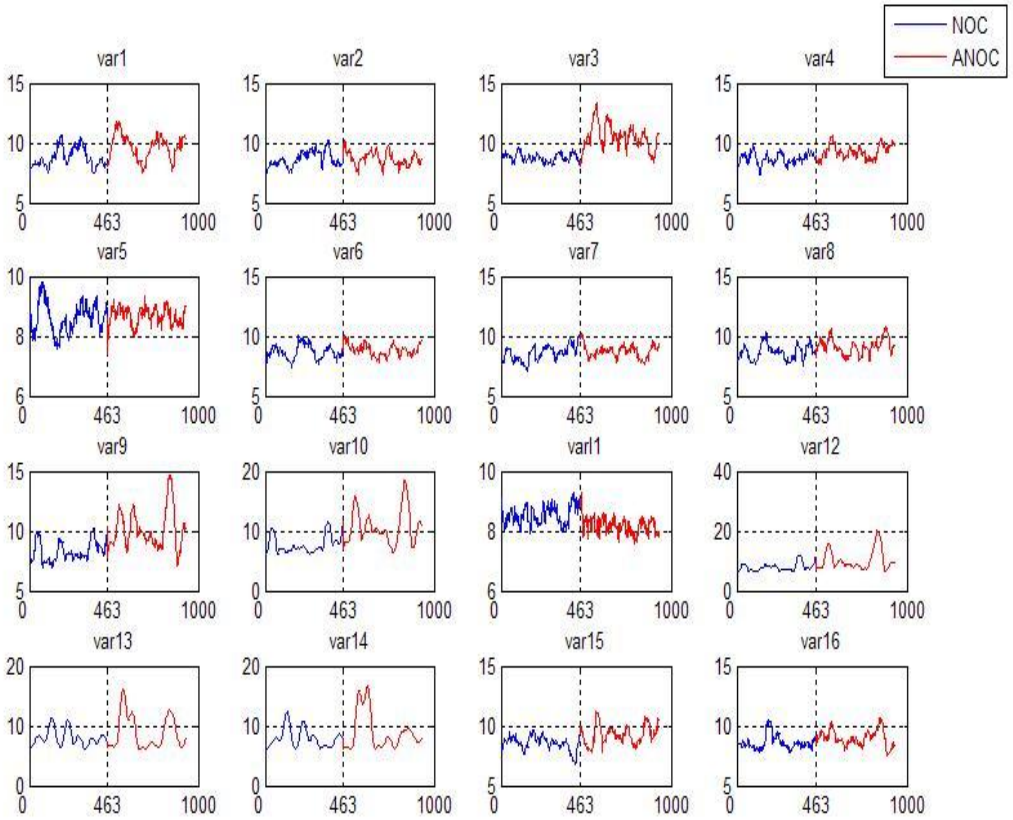


Figure 5.61: Mean value of variables in dissimilarity matrix in normal condition (NOC) (first 463 data samples) and fault condition (ANOC) (last 463 data samples) in Case 9. In each plot the horizontal axis represents the number of observations in the mean value of the dissimilarity matrix of the lagged variables and the vertical axis represents their corresponding mean values.

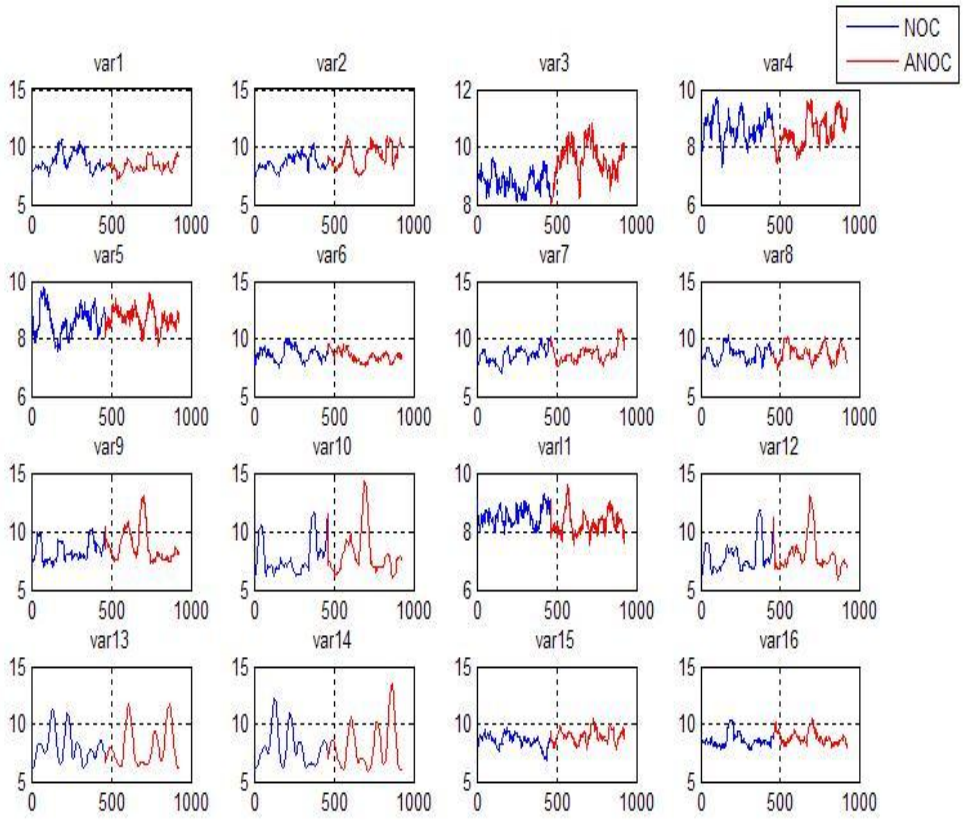


Figure 5.62: Mean value of variables in dissimilarity matrix in normal conditions (NOC) (first 463 data samples) and fault conditions (ANOC) (last 463 data samples) in Case 15. In each plot the horizontal axis represents the number of observations in the mean value of the dissimilarity matrix of the lagged variables and the vertical axis represents their corresponding mean values.

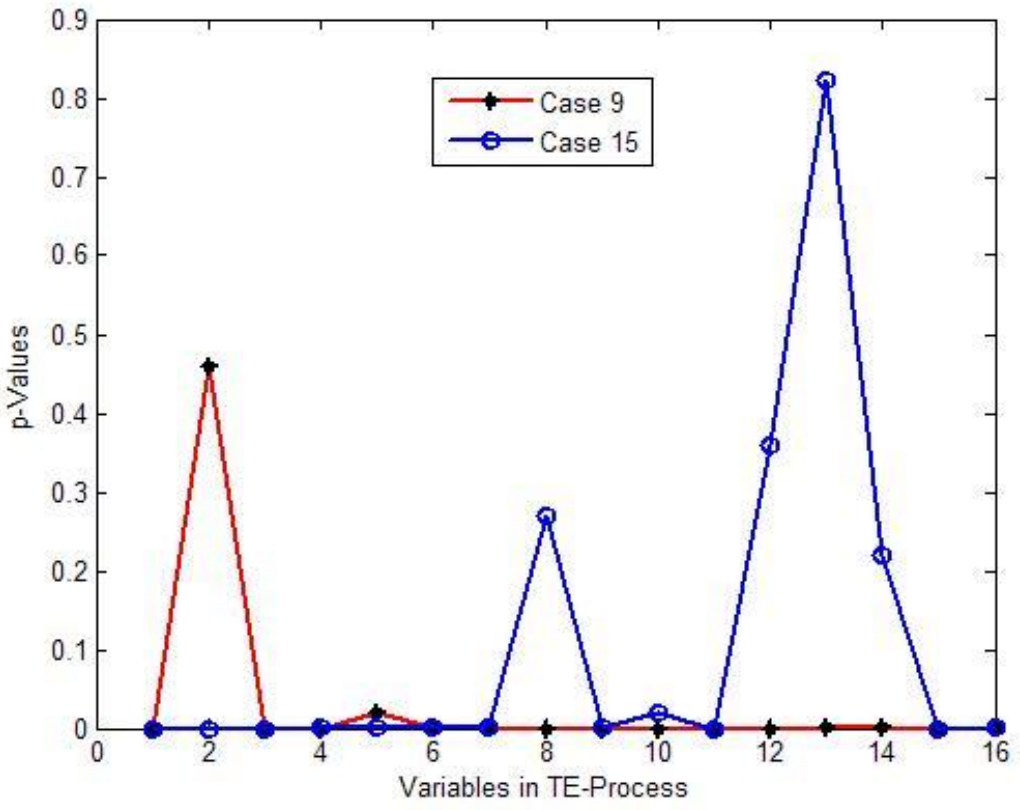


Figure 5.63: The p-values for the paired t-test on the mean values of variables in dissimilarity matrix in normal and fault condition in Case 9 and Case 15.

In general a relatively better performance of multimodal nonlinear SSA was achieved in monitoring all the fault conditions in the TE-Process. This underscores the advantage of using nonlinear SSA in monitoring where the transformed variation in the data can be obtained through the extraction of fewer components (eigenvectors and eigenvalues) for the reconstruction of data into multiple modes. The high percentage of reliability obtained in both $\boldsymbol{T}^2$ and SPE statistics of MM-KDSSA and MM-DSSA showed the efficiency of these methods in capturing all the random variations in the process as well as the nonlinear correlation between the process variables using inter-dissimilarity measure, which naturally incorporate the issue of nonlinearity in the data. The fault detection in MM-KDSSA was achieved with a relatively lower number of modes as compared to MM-DSSA and MS-SSA because of the dissimilarity scale transformation using kernel functions in KDSSA. The MM-NLSSA approach also achieved a reliability

percentage that is higher than or equal to MM-KDSSA and MM-DSSA in most of the fault cases in mode 1 and 2 in this study.

CHAPTER 6: CONCLUSIONS

Three different novel multimodal and multivariate statistical process monitoring techniques, based on nonlinear singular spectrum analysis have been proposed in this study for the early and accurate detection of faults in the nonlinear, dynamic and complex chemical process systems. SSA is a method used for the the analysis of time series structures. The main purpose of SSA is the decomposition of a time series into additive components that can be associated with a trend, oscillatory patterns that are possibly amplitude- or phase-modulated as well as a periodic or noise component. In the three novel nonlinear SSA process monitoring methods, the first approach used the covariance measure between the variables while the second and third approaches used the dissimilarity measures to extract nonlinear latent variables from the chemical process systems.

An important advantage of SSA compared to other methods is its adaptive nature. Specifically, the basis functions used for time series decomposition are obtained from the data itself. This allows for a better and more compact representation of certain features in the data, such as nonlinear harmonics that can be obtained using fixed basis functions such as sinusoids in Fourier analysis or dilated and translated mother wavelets in wavelet analysis. Process monitoring using nonlinear SSA is based on obtaining a multimodal representation of a multivariate time series and, subsequently applying a standard statistical process control scheme to this representation. Moreover, SSA explicitly accounts for the auto-correlation in the process data and can also effectively decompose a short time series into deterministic and stochastic components in multiple modes.

Recently, a variant of SSA known as nonlinear SSA has been proposed for nonlinear extraction of features from the time series by recombining the SSA modes in order to reduce the energy scatter into many SSA modes when the underlying signal has oscillatory components. This motivated the extension of the application of nonlinear SSA to extract nonlinear components in chemical process systems. MSPC methods with nonlinear SSA called multimodal NLSSA, DSSA and KDSSA were developed in this

study for monitoring nonlinear process systems, which use the concept of nonlinear PCA, multidimensional scaling and kernel MDS for the multimodal decomposition of data. The application of multimodal NLSSA, multimodal DSSA and multimodal KDSSA are demonstrated using a simulated nonlinear data source.The monitoring performance of multimodal NLSSA, multimodal DSSA and multimodal KDSSA are compared to conventional PCA and MS-SSA using the Tennessee Eastman process data and data from an industrial source.

In the MM-NLSSA method, the process data is decomposed into multiple modes using SSA and then the nonlinear relations among the PCs of decomposed data in the different modes are extracted using NLPCA.cir. Hence, MM-NLSSA is an alternative data adaptive nonlinear approach to uncover the nonlinear characteristics while simultaneously providing a fault detection platform for nonlinear chemical processes as illustrated by case studies in this thesis. The proposed monitoring strategy extends the suitability of NLSSA to monitor statistically processes based on nonlinear auto-correlated measurements. Additionally, the resulting nonlinear PCA model at multiple levels is more sensitive to detecting changes in a process. These ideas are illustrated by simulated and industrial case studies.

In MM-KDSSA and MM-DSSA, the decomposition step in basic SSA is modified with kernel methods and CMDS techniques for nonlinear and multimodal extraction of components in the data. The difference between KDSSA and DSSA is that KDSSA uses the CMDS in the feature space, which is the realisation of the dissimilarity measures of the points in the trajectory matrices by Gaussian kernels, while the latter uses linear kernels for the transformation of the dissimilarity measures.

In MM-DSSA, the augmented variables are transformed into principal scores of lower dimensions using dissimilarity scales/inter-distance scales of the respective sample points and then the augmented data was reconstructed into multiple modes through diagonal averaging. In other words, the role of PCA in basic SSA decomposition is replaced with CMDS based MSPM framework. MM-DSSA is thus an alternative nonlinear process monitoring technique to resolve the issues with conventional MSPC techniques in chemical process monitoring. The proposed multimodal monitoring strategy broadens the use of dissimilarity scale-based SSA to monitor nonlinear process at multiple modes as illustrated by simulated case study.

In MM-KDSSA, Gaussian kernels were used in an SSA framework to map the input data (inner product) to a higher dimensional space for SSA decomposition with CMDS for the nonlinear extraction of multi-level components. In MM-KDSSA the dissimilarity/inter-distance measures of the augmented data are transformed to a feature space using a nonlinear map. The reconstruction of the augmented data into multiple modes is obtained by the application of CMDS in feature space through diagonal averaging. In other words, the role of PCA in basic SSA decomposition is replaced with a kernel MDS-based MSPM framework. Thus, MM-KDSSA is an alternative nonlinear process monitoring technique to resolve the issues with conventional nonlinear MSPC techniques in chemical process monitoring. The proposed nonlinear monitoring techniques made kernel dissimilarity scale based SSA more appropriate to monitor the nonlinear process at multiple scales, as illustrated by simulated and industrial case studies.

In this study, it has been demonstrated that nonlinear SSA approaches to process monitoring compares favourably with existing methods such as MS-SSA and PCA. The performace of these methods were compared using percentage of samples, which exceeded the control limits of monitoring statistics (percentage reliability) on the simulated data, base metal flotation plant data and the Tennessee Eastman industrial control test problem. Nonlinear SSA could reliabily detect parameter changes in the simulated multivariate nonlinear data. However, nonlinear SSA methods outperformed PCA and MS-SSA in the detection of faults in base metal flotation plant data. When applied to data generated from the Tennessee Eastman control model, nonlinear SSA methods performed better than both PCA and MS-SSA on most of the difficult, unobservable disturbances considered. The overall performance of MM-KDSSA on all case studies can be perceived as superior compared to PCA, MS-SSA and other nonlinear SSA methods. Compared to other nonlinear spectral methods, the main advantage of MM-KDSSA is that it does not involve nonlinear optimisation and handles a wide range of nonlinearities in the data by using kernel functions as illustared in the case studies.

In summary, from the simulated and industrial applications, MM-KDSSA was found to be more efficient and sensitive in detecting various faults in the process data compared to MM-NLSSA and MM-DSSA in relatively smaller modes (first mode) in the Tennessee

Eastman process while other nonlinear methods seemed to be almost equally efficient in detecting parameter changes. This is achieved in MM-NLSSA in a smaller number of modes just as in MM-KDSSA while MM-DSSA required a larger number of modes for reconstruction components to become effective in fault detection. In short, MM-KDSSA and MM-NLSSA can perform better by monitoring reconstructed components in the first two modes. On the other hand, MM-NLSSA requires large data sets for model caliberation for effective fault detection and also has difficulties in deciding the number of mapping and demapping layers when used with an auto-associative neural network.

Further modifications of the current study can be executed to modify the proposed monitoring methods. Firstly, general theoretical guidelines can be used for the optimal selection of window length and selection of significant modes for fault detection. Secondly the effectiveness of the proposed nonlinear SSA methods can be improved by the multiscale decomposition of data using multiscale nonlinear extensions of SSA and varying embedding dimension instead of fixed window size, which was used in this study. Thirdly, process monitoring in MM-KDSSA and MM-DSSA can be further investigated by considering the effect of other types of interdistance scaling such as city-block scale to find dissimilarities between the data points instead of Euclidean distance, which was used in this study.

REFERENCES

Abarbanel, H.D.I. 1997. *Analysis of observed chaotic data*. New-York: Springer-Verlag.

Alcala, C.F. & Qin, S.J. 2010. Reconstruction-based contribution for process monitoring with kernel principal component analysis. *Industrial and Engineering Chemistry Research*, 49, 7849–7857.

Aldrich, C. & Barkhuizen, M. 2003. Process system identification strategies based on the use of singular spectrum analysis. *Minerals Engineering*, 16, 815-826.

Alexandrov, T. 2009. A method of trend extraction using singular spectrum analysis. *Statistical Journal,* 7, 1–22.

Alonso, F.J. & Salgado, D.R. 2005. Applicaion of singular spectrum analysis to tool wear detection using sound signals. *IMechE, Journal of Engineering Manufacture*, 219, 703–710.

Alonso, F.J., Castillo Del, J.M. & Pintado, P. 2005. Application of singular spectrum analysis to the smoothing of raw kinematic signals. *Journal of Biomechanics,* 38, 1085–1092.

Antory, D., Irwin, G.W., Kruger, U. & McCullough, G. 2008. Improved process monitorig using nonlinear principal component models. *International Journal of Intelligent Systems*, 23, 520–544.

Aradhye, H.B., Davis, J.F. & Bakshi, B.R. 2003. Multiscale SPC using wavelet-theoretical analysis and properties. *AIChE Journal*, 49, 939–958.

Bakshi, B.R. 1998. Multiscale PCA with applications to multivariate statistical process monitoring. *AIChE Journal*, 44, 1596–1610.

Barkhuizen, M. & Aldrich, C. 2003. Dynamic process analysis with non-linear Monte Carlo singular spectrum analysis. Proceedings of the First African Control Conference, Cape Town, South Africa.

Borg, I. & Groenen, P. 1997. *Modern multidimensional scaling: Theory and ppplications*. New York: Springer-Verlag.

Bray, M.A. & Wikswo, J.P. 2002. Consideration in phase plane analysis for nonstationary reentrant cardiac behaviour. *Physical Review E*, 65, 1902-1909.

Broomhead, D. & King, G. 1986. Extracting qualitative dynamics from experimental data. *Physica D*, 20, 217–236.

Carniel, R., Barazza, F. & Pascolo, P. 2006. Improvement of nakamura technique by singular spectrum analysis. *Soil Dynamics and Earthquake Engineering*, 26, 55–63.

Chakour, C., Harkat, M.F. & Djeghaba, M. 2015. New adaptive kernel principal component analysis for nonlinear dynamic process monitoring. *Applied Mathematics and Information Sciences*, 9, 1833–1845.

Cheng, C.Y., Hsu, C.C. & Chen, M.C. 2010. Adaptive kernel principal component analysis (KPCA) for monitoring small disturbances of nonlinear processes. *Industrial and Engineering Chemistry Research,* 49, 2254–2262.

Chiang, L., Russell, E. & Braatz, R. 2001. *Fault detection and diagnosis in industrial systems*. London, U.K.: Springer -Verlag.

Choi, J.H., Lee, J.M., Choi, S.W., Lee, D.W. & Lee, I.B. 2005. Fault identification for process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 60, 279–288.

Choi, S.W. & Lee, I.B. 2004. Nonlinear dynamic process monitoring based on dynamic kernel PCA. *Chemical Engineering Science*, 59, 5897–5908

Choi, S.W., Morris, J. & Lee, I. 2008. Nonlinear multiscale modelling for fault detection and identification. *Chemical Engineering Science*, 63, 2252–2266.

Cox, T.F. & Cox, M.A.A. 1994. *Multidimensional scaling*. London: Chapman & Hall.

Cox, T.F. 2001. Multidimensional scaling used in multivariate statistical process control. *Journal of Applied Statistics*, 28: 365–378.

Cox, T.F. (Ed.). 2003. Multidimensional scaling in process control. In Khattree, R. & Rao, C.R. (eds.), *Handbook of Statistics*. Elsevier Science B.V.

Das, A., Maiti, J. & Banerjee, R.N. 2012. Process monitoring and fault detection strategies: a review. *International Journal of Quality & Reliability Management*, 29: 720–752.

Daubechies, I. 1992. *Ten lectures in wavelets, SIAM,* Philadelphia, P.A.

Deng, X. & Tian, X. 2006. Multivariate statistical process monitoring using multi-scale kernel principal component analysis. Supervision and Safety of Technical Processes, IFAC Fault Detection, Beijing.

Dong, D.L. & McAvoy, T.J 1996. Nonlinear principal component analysis - based on principal curves and neural networks. *Computers and Chemical Engineering*, 20, 65–78.

Downs, J.J. & Vogel, E.F. 1993. A plant-wide industrial process control problem. *Computers and Chemical Engineering*, 17, 245–255.

Doymaz, F., Chen,J., Romagnoli,J.A. & Palazoglu, A. 2001. A robust strategy for real time process monitoring. *Journal of Process Control,* 11, 343–359.

Elsner, J. & Tsonis, A. 1996. Singular spectrum analysis–a new tool in time series analysis. New York, N.Y.: Plenum Press.

Eslamloueyan, R. 2011. Designing a hierarchical neural network based on fuzzy clustering for fault diagnosis of the Tennessee-Eastman process. *Applied Soft Computing*, 11, 1407–1415.

Florinsky, I., Eilers, R., Wiebe, B. & Fitzgerald, M. 2009. Dynamics of soil salinity in the Canadian prairies: Application of singular spectrum analysis. *Environmental Modelling & Software*, 24, 1182–1195.

Fourie, S.H. & De Vaal, P. 2000. Advanced process monitoring using an online non-linear multiscale principal component analysis methodology. *Computers and Chemical Engineering*, 24, 755–760.

Frank, P.M. 1990. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—a survey and some new results. *Automatica*, 26, 459–474.

Ganesan, R., Das, T.K. & Venkataramanan, V. 2004. Wavelet-based multiscale process monitoring: A literature review. *IEEE Transactions*, 36, 787–806.

Ge, Z., Song, Z& Gao, F. 2013. Review of recent research on data-based process monitoring. *Industrial and Engineering Chemistry Research*, 52, 3543–3562.

Ge, Z.Q., Yang, C.J. & Song, Z.H. 2009. Improved kernel PCA-based monitoring approach for nonlinear processes. *Chemical Engineering Science*, 64, 2245–2255.

Geng, Z. & Zhu, Q. 2005. Multiscale nonlinear principal component analysis (NLPCA) and its application for chemical process monitoring. *Industrial and Engineering Chemistry Research*, 44, 3585–3593.

Ghil, M. & Yiou, P. 1996. Spectral methods: What they can and cannot do for climatic time series. In Anderson, D. & Willebrand, J. (eds.), *Decadal climatic variability. Dynamics and predicability*. Amsterdam: Elsevier, 445–482.

Ghil, M., Allen, M., Dettinger, M., Ide, K., Kondrashow, D., Mann, M., Robertson, A., Sauders, A., Tian, Y., Varadi, F. & Yiou, P. 2002. Advanced spectral methods for cimatic time series. *Reviews of Geophysics*, 40, 3.1–3.41.

Golyandina, N., Nekrutkin, V. & Zhigljavsky, A. 2001. Analysis of time series structure—SSA and related techniques. Boca Raton, Florida: Chapman & Hall/CRC Press.

Golyandina, N. & Shlemov, A. 2014. Variations of singular spectrum analysis for separability improvement: Non-orthogonal decompositions of time series. *Statistics and Its Interface*. 1-15.

Harkat, M., Djelel, S., Doghmane, N. & Benouaret, M. 2007. Sensor fault detection, isolation and reconstruction using nonlinear principal component analysis. *International Journal of Automation and Computing,* 4, 149-155.

Hassani, H. 2007. Singular specrum analysis: Methodology and comparison. *Journal of Data Science*, 5, 239–257.

Hassani, H., Heravi, A. & Zhigljasky, A. 2009. Forecasting European industrial production with singular spectrum analysis. *International Journal of Forecasting*, 25, 103–118.

Hassani, H. & Zhigljavsky, A. 2009. Singular spectrum analysis: Methodology and application to economic data. *Journal of Systems Science and Complexity,* 22, 372–394.

Hassani, H., Soofi, A. & Zhigljavsky, A. 2010. Predicting daily exchange rate with singular spectrum analysis. *Nonlinear Analysis: Real World Applications*, 11, 2023–2034.

Hassani, H. & Thomakos, D. 2010. A review on singular spectrum analysis for economic and financial time series. *Statistics and Its Interface,* 3, 377–397.

Hassani, H., Heravi, A. & Zhigljasky, A. 2013. Forecasting UK industrial production with multivariate singular spectrum analysis. *Journal of Forecasting*, 32, 395–408.

Hastie, T. & Stuetzle, W. 1989. Principal curves. *Journal of the American Statistical Association*, 84, 502–516.

Hastie, T., Tibshirani, R. & Friedman, J. 2011. The elements of statistical learning: Data mining. *Inference and Prediction*.

Himmelblau, D.M. 1978. *Fault detection and diagnosis in chemical and petrochemical processes*. New York: Elsevier Scientific Publishing Co.

Hotelling, H. 1947. *Multivariate quality control*. New York: McGraw-Hill.

Hsieh, W.W. 2001. Nonlinear principal component analysis by neural networks. *Tellus*, 53(A), 599–615.

Hsieh, W.W. & Wu, A. 2002. Nonlinear multichannel singular spectrum analysis of the tropical Pacific climate variability using a neural network approach. *Journal of Geophysical Research*, 107, 13-1-13-15.

Hsieh, W.W. & Hamilton, K. 2003. Nonlinear singular spectrum analysis of the tropical stratospheric wind. *Quarterly Journal of the Royal Meteorological Society*, 129, 2367–2382.

Hsieh, W.W. 2004. Nonlinear multivariate and time series analysis by neural network methods. *Review of Geophysics*, 42, RG1003.

Hsieh, W.W. 2007. Nonlinear principal component analysis by neural networks. *Neural Networks*, 20, 434–443.

Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Yen, N.C., Tung, C. & Liu, H. 1998. The emperical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London*, Series A 454, 903–995.

Jackson, J.E. 1991. *A user's guide to principal components*. New York: Wiley.

Jemwa, G.T. & Aldrich, C. 2006. Classification of process dynamics with Monte Carlo singular spectrum analysis. *Computers and Chemical Engineering*, 30, 816–831.

Jia, F., Matin, E.B. & Morris, A.J. 2001. Nonlinear principal component analysis with applications to process fault detection. *International Journal of Systems Science*, 31, 1473–1487.

Kano, M., Ohno, H., Hasebe, S. & Hashimoto, I. 2001. A new multivariate statistical process monitoring method using principal compnent analysis. *Computers and Chemical Engineering*, 25, 1103–1113.

Kano, M., Hasebe, S., Hashimoto, I. & Ohino, H. 2002. Statistical process monitoring based on dissimilarity of process data. *AIChE Journal*, 48, 1231–1240.

Kano, M., Tanaka, S., Hasebe, S., Hashimoto, I. & Ohino, H. 2003. Monitoring independent compnents for fault detection. *AIChE Journal*, 49, 969-976.

Kano, M. & Nakagawa, Y. 2008. Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry. *Computers and Chemical Engineering,* 32, 12–24.

Karpenko, M., Sepehri, N. & Scuse, D. 2003. Diagnosis of process valve actuator faults using a multilayer neural network. *Control Engineering Practice*, 11, 1289–1299.

Kim, B., Kim, J. & Choi, S. 2009. Use of neural network to model X-ray photoelectron spectroscopy data for diagnosis of plasma etch equipment. *Expert Systems with Applications,* 36, 11347–11351.

Kirby, M.J. & Miranda, R. 1996. Circular nodes in neural networks. *Neural Computation*, 8, 390–402.

Kourti, T., Lee, J. MacGregor, J.F. 1996. Experiences with industrial applications of projection methods for multivariate statistical process control. *Computers and Chemical Engineering*, 20, 745–750.

Kramer, M.A. 1991. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 32, 233–243.

Kresta, J.V., MacGregor, J.F. & Martin, T.F. 1991. Multivariate statistical monitoring of process operating performance. *The Canadian Journal of Chemical Engineering*, 69, 35–47.

Krishnannair, S. 2010. *Multiscale process monitoring with singular spectrum analysis*. MScEng thesis. Stellenbosch: Stellenbosch University.

Kruger, U., Antory, D., Hahn, J., Irwin, G.W. & McCullough, G. 2005. Introduction of a nonlinearity measure for principal component models. *Computers and Chemical Engineering,* 29, 2355–2362.

Kruskal, J.B. & Wish, M. 1978. *Multidimensional scaling*. California, USA: SAGE Publications.

Ku, W., Storer, R.H. & Geogakis, C. 1995. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems,* 30, 179–196.

Lau, C.K., Ghosh, K., Hussain, M.A. & Che Hassan, C.R. 2013. Fault diagnosis of Tenessee Eastman process with multi-scale PCA and ANFIS. *Chemometrics and Intelligent Laboratory Systems*, 120, 1–14.

Lee, D.S., Park, J.M. & Vanrolleghem, P.A. 2005. Adaptive multiscale principal component analysis for on-line monitoring of a sequencing batch reactor. *Journal of Biotechnology*, 116, 195–210.

Lee, D.S., Lee, M.W., Woo, H., Kim, Y.J. & Park, J.M. 2006. Multivariate online monitoring of a full-scale biological anaerobic filter process using kernel-based algorithm. *Industrial and Engineering Chemistry Research,* 45, 4335–4344.

Lee, J.M., Yoo, C.K., Choi, S.W., Vanrolleghem, P.A. & Lee, I.B. 2004. Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 60, 223–234.

Lee, J.M., Qin, S.J. & Lee, I.B. 2007. Fault detection of non-linear process using kernel independent component analysis. *Canadian Journal of Chemical Engineering*, 2007, 526–536.

Li, H. & Xiao, D. 2011. Fault diagnosis of Tennessee Eastman process using signal geometry matching technique. *EURASIP Journal on Advances in Signal Processing*, 83, 1–19.

Lin, W., Qian, Y. & Li, X. 2000. Non-linear dynamic principal component analysis for on-line process monitoring and diagnosis. *Computers and Chemical Engineering*, 24, 423–429.

Lowry, C.A., Woodall, W.H., Champ, C.W. & Rigdon, S.E. 1992. A multivariate exponentially weighted moving average chart. *Technometrics,* 34, 46–53.

Luo, R., Misra, M. & Himmelblan, D.M. 1999. Sensor fault detection via multiscale analysis and dynamic PCA. *Industrial and Engineering Chemistry Research*, 38, 1489–1495.

MacGregor, J.F., Jaekle, C., Kiparissides, C. & Koutoudi, C. 1994. Process monitoring and diagnosis by multiblock PLS methods. *AIChE Journal,* 40, 826–838.

MacGregor, J.F. & Kourti, T. 1995. Statistical process control of multivariate processes. *Control Engineering Practice*, 3, 403–414.

Mamou, J. & Feleppa, E.J. 2007. Singular spectrum analysis applied to ultrasonic detection and imaging of brachytherapy seeds. *Journal of Acoustical Sciety of America,* 121, 1790–1801.

Maulud, A., Wang, D. & Romagnoli, J.A. 2006. A multi-scale orthogonal nonlinear strategy for multi-variate statistical process monitoring. *Journal of Process Control,* 16, 671–683.

Meyer, Y. 1992. Wavelets and operators. New York & Cambridge: Cambridge University Press.

Mineva, A. & Popivanov, D. 1996. Method for single-trial readiness potential identification, based on singular spectrum analysis. *Journal of Neuroscience Methods*, 68, 91–99.

Mirin, S.N.S. & Wahab, N.A. 2014. Fault detection and monitoring using multiscale principal component analysis at sewage treatment plant. *Journal Teknologi*, 70, 87–92.

Misra, M.Y.H., Qin, S.J. & Ling, C. 2002. Multivariate process monitoring and fault diagnosis by multiscale PCA. *Computers and Chemical Engineering*, 26, 1281–1293.

Mtheus, J., Dourado, A., Henriques, J., Antonio, M. & Nogueira, D. 2006. Iterative multidimensional scaling for industrial process monitoring. *IEEE International Conference on Systems, Man, and Cybernetics*. Taipei, Taiwan.

Negiz, A. & Cinar, A. 1997. Statistical monitoring of multivariate dynamic processes with state-space models. *AIChE Journal*, 43, 2002–2012.

Nomikos, P. & MacGregor, J.F. 1995. Multivariate SPC charts for monitoring batch processes. *Technometrics*, 37, 41–59.

Oropeza, V. & Sacchi, M. 2011. Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis. *Geophysics*, 76, V25-V32.

Palomo, M.J., Sanchis, R., Verdu, G. & Ginestar, D. 2003. Analysis of pressure signals using a singular spectrum analysis (SSA) methodology. *Progress in Nuclear Energy*, 43, 329–336.

Pekalska, E., Paclik, P. & Duin, R.P.W. 2001. A generalized kernel approach to dissimilarity-based classification. *Journal of Machine Learning Research*, 2, 175–211.

Pereira, W.C. & Maciel, C.D. 2001. Performance of ultrasound echo decomposition using singular spectrum analysis. *Ultrasound in Medicine and Biology*, 27, 1231–1238.

Pereira, W.C., Bridal, S.L., Coron, A. & Laugier, P. 2004. Singular spectrum analysis applied to backscattered ultrasound signal from in vitro human cacerous bone specimens. *IEEE Transactions on Ultrason Ferroelectrics Frequency Control*, 51, 302–312.

Polikar, R. 1996. The Wavelet tutorial part III. Multiresolution analysis & the continuous wavelet transform. [Online] Available: https://ccrm.stanford.edu/WTpart/html Accessed: 18 August 2009.

Prentice, I.C. 1980. Multidimensional scaling as a research tool in quanternary palynology: A review of theory and methods. *Review of Palaeobotany & Palynology*, 31, 71–104.

Qin, S.J. 2003. Statistical process monitoring: Basics and beyond. *Journal of Chemometrics*, 17, 480–502.

Raich, A. & Cinar, A. 1996. Statistical process monitoring and disturbance diagnosis in multivariate continuous processes. *AIChE Journal*, 42, 995–1009.

Ralston, P., DePuy, G. & Graham, J.H. 2004. Graphical enhancement to support PCA-based process monitoring and fault diagnosis. *ISA Transactions*, 43, 639–653.

Reis, M., Saraiva, P. & Bakshi, B. 2008. Multiscale statistical process control using wavelet packets. *AIChE Journal*, 54, 2366–2378.

Romdhani, S., Gong, S. & Psarrou, A. 1999. *A multi-view nonlinear active shape model using kernel PCA*. Proceedings of BMVC, Nottingham, UK.

Rosen, C. & Lennox, J.A. 2001. Multivariate and multiscale monitoring of waste water treatment operation. *Water Research*, 35, 3402–3410.

Rotem, Y., Wachs, A. & Lewin, D.R. 2000. Ethylene compressor monitoring using model-based PCA. *AIChE Journal*, 46, 1825–1836.

Rozynski, G., Larson, M. & Pruszak, Z. 2001. Forced and self-organized shoreline response for a beach in the southern Baltic Sea determined through singular spectrum analysis. *Coastal Engineering,* 43, 41–58.

Russell, E.L., Chiang, L.H. & Braatz, R.D. 2000. *Data-driven techniques for fault detection and diagnosis in chemical processes*. London: Springer.

Salgado, D.R. & Alonso, F.J. 2006. Tool wear detection in turning operations using singular spectrum analysis. *Journal of Materials Processing Technology*, 171, 451–458.

Salgado, D.R. & Alonso, F.J. 2007. An approach based on current and sound signals for i-process tool wear monitoring. *International Journal of Machine Tools & Manufacture*, 47, 2140–2152.

Sanger, T.D. 1989. Optimal unsupervised learning in a single-layer linear feedforward network. *Neural Networks*, 2, 459–473.

Schoellhamer, D.H. 2001. Singular spectrum analysis for time series with missing data. *Geophysicl Research Letters*, 28, 3187–3190.

Schölkopf, B. 2000. The kernel trick for distances. *Advances In Neural Information Processing Systems*. Cambridge, UK.

Schölkopf, B., Smola, A. & Muller, K. 1998. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10, 1299–1399.

Shams, M.B., Budman, H. & Duever, T. 2010. Fault detection using CUSUM based techniques with application to the Tennessee Eastman Process. *Proceedings of the*

9th International Symposium on Dynamics and Control of Process Systems, Leuven, Belgium, Mayuresh Kothare, Moses Tade, Alain Vande Wouwer, IIse Smets

Shao, R., Jia, F., Martin, E.B. & Morris, A.J. 1999. Wavelets and non-linear principal components analysis for process monitoring. *Control Engineering Practice*, 7, 865–879.

Silva, R.G. 2010. Condition monitoring of the cutting process using a self-organizing spiking neural network map. *Journal of Intelligent Manufacturing*, 21, 823–829.

Strang, G. 2009. *Introduction to linear algebra*. Wellesley: Cambridge University Press.

Sun, J., Chai, Y., Guo, M. & Li, H.F. 2012. On-line adaptive principal component extraction algorithm using iteration approach. *Signal Processing,* 92, 1044–1068.

Takane, Y. 2003. Matrices with special reference to applications in psychometrics. *Linear Algebra and its Applications,* 388, 341–361.

Tan, S. & Mavrovouniotis, M.L. 1995. Reducing data dimensionality through optimizing neural network inputs. *AIChE Journal*, 41, 10.

Tan, S.C., Lim, C.P. & Rao, M.V.C. 2007. A hybrid neural network model for rule generation and its application to process fault detection and diagnosis. *Engineering Applications of Artificial Intelligence*, 20, 203–213.

Thomakos, D.D., Wang, T. Wille, L.T. 2002. Modeling daily realized futures voltality with singular spectrum analysis. *Physica,* 312, 505–519.

Tzagkarakis, G., Papadopouli, M. & Tsakalides, P. 2007. Singular spectrum analysis of traffic workload in a large-scale wireless LAN. MSWIM'07. Chania, Crete Island, Greece.

Van der Maaten, L.J.P., Postma, E.O. & Van den Herik, H.J. 2008. Dimensionality reduction: A comparative review. *Review: Literature and Arts of the Americas*, 10, 1–35.

Vautard, R. & Ghil, M. 1989. Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D: Nonlinear Phenomina,* 35, 395–424.

Vautard, R., Yiou, P. & Ghil, M. 1992. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D*, 58, 95–126.

Vitanov, N.K., Sakai, K. & Dimitrova, Z.I. 2008. SSA, PCA, TDPSC, ACFA: Useful combination of methods for analysis of short and nonstationary time series. *Chaos, Solitons and Fractals*, 37, 187–202.

Wang, W.J., Chen, J. & Wu, X.K. 2001. The application of some non-linear methods in rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 15, 697–705.

Williams, C.K.I. 2002. On a connection between kernel PCA and metric multidimensional scaling. *Machine Learning*, 46, 11–19.

Wise, B.M. & Gallagher, N.B. 1996. The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control,* 6, 329–348.

Wold, S., Esbensen, K. & Geladi, P. 1987. Principal component analysis. Chemometrics and Intelligent Laboratory Systems, 2, 37–52.

Wold, S. 1994. Exponentially weighted moving principal component analysis and projection to latent structures. *Chemometrics and Intelligent Laboratory Systems*, 23, 149–161.

Wold, S., Kettane, H.N. & Tjessem, K. 1996. Hierarchical multiblock PLS and PC model for easier model interpretation and as an alternative to variable selection. *Journal of Chemometrics*, 10, 463–482.

Wu, B. & Yu, J.B. 2010. A neural network ensemble model for on-line monitoring of process mean and variance shifts in correlated processes. *Expert Systems with Applications*, 37, 4058–4065.

Wu, H. & Gong, J. 2000. Forecast of network behavior based on singular spectrum analysis. *IEEE Proceedings,* 1, 702–706.

Xu, L.O., E., Suen, C. 1992. Modified hebbian learning for curve and surface fitting. *Neural Networks: the joint official journal of International Neural Network Society*, 5, 441–457.

Yang, M., Yao, T., Wangh, H. & Gou, X. 2006. Climatic oscillations over the past 120kyr recorded in the Guliya ice core, China. *Quaternary International,* 154–155, 11–18.

Yin, S., Ding, S.X., Xie, X. & Luo, H. 2014. A review on basic data-driven approaches for industrial process monitoring. *IEEE Transactions on Industrial Electronics*, 61, 6418–6428.

Yiou, S., Sonette, D. & Ghil, M. 2000. Data-adaptive wavelets and multiscale singular spectrum analysis. *Physica D,* 142, 254–290.

Yoon, S. & MacGregor, J.F. 2004. Principal component analysis of multiscale data for process monitoring and fault diagnosis. *AIChE Journal*, 50, 2891–2903.

Yuan, H. 2015. *Process analysis and performance assessment for sheet forming processes*. PhD dissertation. Kingston, Ontario: Queen's University.

Yunus, M.Y.M. & Zhang, J. 2010. A multidimensional scaling based process montoring technique. 20th *European Symposium on Computer Aided Process Engineering (ESCAPE20).* Elsevier B.V.

Yunus, M.Y.M. & Zhang, J. 2010. Multivariate process monitoring using classical multidimensional scaling and procrustes analysis. 9th *International Symposium on Dynamics and Control of Process Systems (DYCOPS).* Mayuresh Kothare, Moses Tade, Alain Vande Wouwer, llse Smets (Eds).

Yunus, M.Y.M. & Zhang, J. 2014. Utilizing of dissimilarity scale-based PCA in mulivariate statistical process monitoring application. *Journal of Applied Sciences*, 14, 1244–1251.

Zhang, J., Martin, E.B. & Morris, A.J. 1997. Process monitoring using non-linear statistical techniques. *Chemical Engineering Journal*, 67, 181–189.

Zhang, Y.W. 2008. Fault detection and diagnosis of nonlinear processes using improved kernel independent component analysis (KICA) and support vector machine (SVM). *Industrial and Engineering Chemistry Research*, 47, 6961–6971.

Zhang, Y.W., Zhou, H., Qin, S.J. & Chai, T.Y. 2010. Decentralized fault diagnosis of large-scale processes using multiblock kernel partial least squares. *IEEE Transactions on Industrial Information*, 6(1), 3–10.

Zhang, Y.W. & Ma, C. 2011. Fault diagnosis of nonlinear processes using multiscale KPCA and multiscale KPLS. *Chemical Engineering Science,* 66, 64–72.

Zhang, Z. & Jordan, M.I. 2009. *Latent variable models for dimensionality reduction*. 12th International Conference on Artificial Intelligence and Statistics (AISTATS), USA.

Zhao, S. & Xu, Y. 2005. Multivariate statistical process monitoring using robust nonlinear principal component analysis. *Tsinghua Science and Technology*, 10, 582–586.